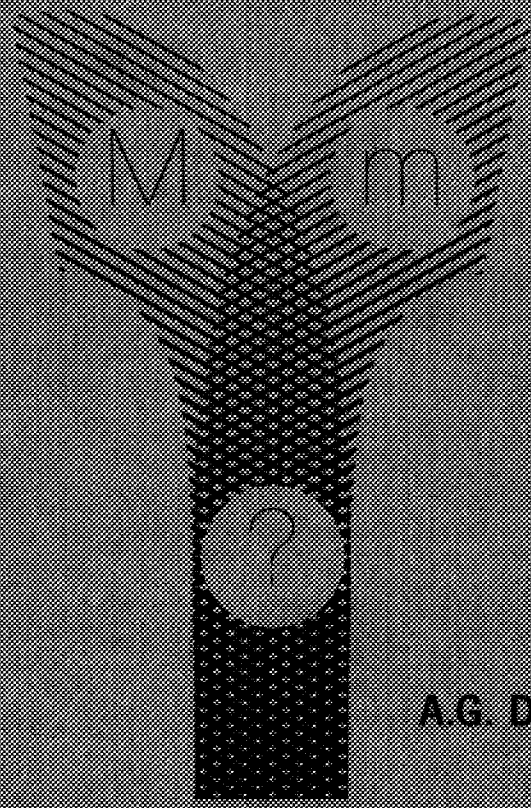
Production-inventory control models

de Kok, A.G.

1985

PRODUCTION-INVENTORY CONTROL MODELS; APPROXIMATIONS AND ALGORITHMS



A.G. DE KOK

VRIJE UNIVERSITEIT TE AMSTERDAM

PRODUCTION-INVENTORY CONTROL MODELS; APPROXIMATIONS AND ALGORITHMS

ACADEMISCH PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN
DE WISKUNDE EN NATUURWETENSCHAPPEN AAN
DE VRIJE UNIVERSITEIT TE AMSTERDAM,
OP GEZAG VAN DE RECTOR MAGNIFICUS
DR. P.J.D. DRENTH,
HOOGLERAAR IN DE FACULTEIT DER SOCIALE WETENSCHAPPEN,
IN HET OPENBAAR TE VERDEDIGEN
OP VRIJDAG 29 MAART 1985 TE 15.30 UUR
IN HET HOOFDGEBOUW DER UNIVERSITEIT, DE BOELELAAN 1105

DOOR

ANTONIUS GERLACUS DE KOK

GEBOREN TE 'S-GRAVENHAGE

1985
CENTRUM VOOR WISKUNDE EN INFORMATICA

PROMOTOR: PROF.DR. H.C. TIJMS

REFERENT: PROF.DR. A. HORDIJK

Aan mijn ouders
Aan Irene

DANKWOORD.

Het voor U liggende proefschrift bevat de weerslag van 4 jaar onderzoek aan de Interfaculteit der Actuariële Wetenschappen en Econometrie. Ik dank allen die gedurende die tijd werkzaam waren aan de Interfaculteit voor de prettige sfeer waarin ik heb kunnen werken.

Velen hebben bijgedragen tot de totstandkoming van dit werk. Een aantal wil ik met name noemen.

Mijn promotor Henk Tijms was degene die de aanzet gaf tot de bestudering van produktie-voorraadmodellen. Zijn enthousiasme heeft mij enorm gestimuleerd. Menig resultaat is tot stand gekomen door gezamenlijk onderzoek. Bij het schrijven van het proefschrift heb ik dankbaar gebruik gemaakt van zijn talloze constructieve aanwijzingen. Gelukkig heeft de gesel van het rode potlood geen striemen nagelaten.

Frank van der Duyn Schouten en Michiel van Hoorn wil ik danken voor de vele discussies en de tijd die ze voor mij hebben vrijgemaakt. Ik ben Arie Hordijk erkentelijk dat hij als referent heeft willen optreden. Tijdens mijn studie in Leiden waren het zijn colleges die mij hebben doen besluiten onderzoek te gaan doen in de mathematische besliskunde.

Gloria Wirz-Wagenaar dank ik voor het typen van het manuscript. Jan-Kees van Ommeren heeft nauwgezet de eerste getypte versie doorgelezen en gecorrigeerd. René Peters en Hans Lugtigheid schreven simulatie-programma's. Marc Salomon heeft de plaatjes gemaakt. Inge Suasso de Lima de Prado-Van Hagen en Mary-Lou Ruhe hebben respectievelijk de Nederlandstalige samenvatting en de Engelstalige inleiding gecorrigeerd. Het omslagontwerp en de finishing touch zijn van R.T. Baanders. Het proefschrift is gedrukt door D. Zwarst en zijn medewerkers van het Centrum voor Wiskunde en Informatica.

Hierbij wil ik mijn ouders danken voor hun steun en aanmoediging bij alles wat ik gedaan heb. Tijdens de 4 jaar van onderzoek was Irene Suasso de Lima de Prado mijn klankbord en baken. Door haar ben ik me er steeds bewust van geweest dat de werkelijkheid meer is dan produktie-voorraadproblemen alleen. Tot slot dank ik Sheila voor al die hartverwarmende begroetingen bij mijn thuiskomst.

CONTENTS

0. INTRODUCTION.

This monograph concerns the probabilistic analysis of a variety of one-product production-inventory models in which the central problem is to coordinate the production rate with the inventory level in order to cope with random fluctuations in demand. Here the main goal is to meet service level constraints corresponding to service measures such as the average number of stockouts per unit time and the long-run fraction of demand to be met directly from stock on hand, while keeping an appropriate balance between the average on-hand inventory and the frequency of changes in the production rate. Our analysis will be guided by the desire to obtain tractable results that are suited for use in practice.

In achieving this goal, we rely heavily upon random walk theory and asymptotic methods from renewal theory as given in Feller [1971].

The control of inventories is one of the major problems in today's industry. Since inventories tie up capital and require storage space one typically wishes to keep inventories low. On the other hand, because of the random nature of the demand process for the product, low inventories will increase the probability of stockout occurrences. Shortages will involve costs as well as loss of goodwill. In practical inventory applications a suitable compromise must be sought between the conflicting alternatives involved when controlling the production rate and inventory level.

In choosing a control rule for replenishing inventories one often tries to minimize certain costs. The costs considered consist of costs of ordering and receiving supplies, costs of holding stocks, costs of manufacturing stocks and costs of running out of stock. The first three types of costs can often be specified. Unfortunately, in many practical situations it is hardly possible to specify the costs of running out of stock. How can the loss of goodwill be quantified? What costs should be associated with future losses and decrease in business because of customers being rejected now? In practice the stockout costs are often introduced indirectly by the use of some service level constraint. Then the corresponding service measure should reflect the manner in which the shortage costs are incurred. For instance, when the shortage costs are proportional to the demand not being met a proper service measure is the fraction of demand that is met directly from stock on hand. When fixed costs are incurred each time demand is not met directly from stock on hand

the average number of unsatisfied demands per unit time might be an appropriate service measure; cf. also Schneider [1981].

There exists an extensive literature on pure inventory models, see e.g. Hadley and Whitin [1963], and Peterson and Silver [1979]. In pure inventory models the main questions to be answered are "when to order" and "how much to order". The replenishment order is received instantaneously or after some lead time. It is important to point out that in pure inventory models the inventory replenishments occur in batches at discrete points in time, whereas in the production-inventory models dealt with in this monograph the inventory replenishments are occurring *continually*. The literature on the latter models is rather limited. The analysis of the production-inventory models with continuous production is usually more intricate than the analysis of pure inventory models.

In the production-inventory models a control rule specifies the production rate at any point in time. A production rate larger than the average demand rate will cause a net increase of the on-hand inventory, and thus may induce high holding costs. A production rate smaller than the average demand rate will cause a net decrease of the on-hand inventory, and thus may induce high shortage costs. An easily implementable control rule achieving a suitable compromise between these two extremes is the so-called (m,M)-rule. Assuming that there are two possible production rates (slow and fast), an (m,M)-rule operates as follows. The production rate is switched from the high value to the low value as soon as the inventory level is at least M, and the production rate is switched back to the high value as soon as the inventory level is less than m.

An important characteristic of the production-inventory control system is the way excess demand is handled. There are two extreme procedures. In the *lost-sales* case any demand in excess of current inventory is lost, whereas in the *backlog* case any demand is backordered until inventory becomes available by production. In practice a combination of these two extreme cases is sometimes used. It will be seen in this monograph that, as opposed to pure inventory models (cf. Tijms and Groenevelt [1984]), for production-inventory problems with continuous production the lost-sales model and the backlog model are essentially different models.

A first attempt to analyse production-inventory models controlled by an (m,M)-rule was made by Gaver [1961], who considered the special model in which the demand process is a compound Poisson process with

exponentially distributed demand and the production is either on or off. For the particular (m,M)-rule with m=0, he derives explicit expressions for several measures of system performance including the long-run average costs. His analysis is based on results from queueing theory. The results of Gaver [1961] were extended considerably in Doshi et al [1978] who presented a renewal-theoretic analysis of Gaver's model for an arbitrary (m,M)-rule, generally distributed demand sizes, and two possible production rates where the slow production rate is not necessarily zero. However, the results obtained in Doshi et al [1978] are computationally tractable only for the special case of exponential demands. For this particular case similar results to those in Doshi et al were obtained by Graves and Keilson [1981] by using a quite different approach. Related results on perishable inventories can be found in Graves [1982], who actually deals with a lost-sales production-inventory model controlled by an (m,M)-rule with m=M, where the demand distribution is either exponential or deterministic. In De Leve et al [1976] a Markov decision method is described for obtaining an average-cost optimal policy for a lost-sales production-inventory model with compound Poisson demand and several production rates. We also mention here the work of Gavish and Graves [1980] and of Tijms [1980], which deals with production-inventory models with a Poisson demand process, and where the items are produced one at a time rather than continuously, cf. also Sobel [1970] for a proof of the optimality of an (m,M)-rule for these models.

In most of the references above the analysis concerns the minimization of the long-run average costs per unit time when assuming a cost structure consisting of fixed costs for switching from one production rate to another and linear holding and shortage costs. The objective of the long-run average costs is also used in the studies of Bather [1966], Doshi [1978] and Vickson [1982], who assume that the demand process is described by a diffusion process rather than by a compound Poisson process. These studies deal not only with the computation of the average costs of a given (m,M)-rule, but also address the question whether an (m,M)-rule is average-cost optimal among all possible control rules.

This monograph studies a variety of production-inventory models with a compound Poisson demand process and two possible production rates, and distinguishes from earlier studies by concentrating on service measures rather than on costs. For a wide class of production-inventory models it will be shown that for the case of generally distributed demand sizes tractable results may be obtained by using fundamental results from

random walk theory and asymptotic methods from renewal theory. The lack of memory of the Poisson process generating the demand epochs runs through the analysis like a continuous thread.

Roughly sketched, our approach is as follows. Firstly, we derive tractable expressions for the service levels associated with the service measures under a given (m,M)-rule. Secondly, we determine the "order quantity" M-m by using holding and switching costs considerations only. Thirdly, we determine the switching level m by invoking the service level requirement.

The sequential approach of separately determining the "order quantity" M-m and the "order-level" m is often followed in practice. For justification of this approach in pure inventory models, see e.g. Peterson and Silver [1979] and Tijms [1986]. These references show that the service level requirement (or the shortage costs) influence the order quantity only to a slight degree in most practical cases. Moreover, these references indicate that choosing the order quantity equal to the well-known economic lot size formula yields a policy that is only short of the optimum in costs. It is clear that a sequential approach of determining M-m and m dramatically reduces the computational effort. So far little attention has been paid to the validity of such an approach for production-inventory problems. We will show that the "economic production quantity" resulting from a deterministic equivalent of our model is in general a good choice for M-m with respect to the minimization of holding and switching costs. In addition we will obtain an improvement of this economic production quantity. It will be seen that this improved quantity leads to an approximately average-cost optimal rule, and is independent of the service level requirement provided the required service level is sufficiently high. Moreover, this quantity turns out to be the same for each of several commonly used service measures. If shortage costs can be specified, and these are linear in one of the service measures considered, then the same quantity is approximately optimal, especially when the shortage costs are large.

The expressions derived for the service levels are in general approximations, since tractable exact results can only be obtained for special cases. Much effort is put into the validation of these approximations. An exhaustive numerical study is performed to indicate where and under what circumstances the various approximations are accurate, or to elucidate the difficulties to which one might be led by uncritical

use of the approximations. Computer simulation is used to validate the approximations. The numerical study yields some rules of thumb for the application of the approximations, including some conclusions about applicability of asymptotic results from renewal theory, which conclusions are of general interest. We find that the approximations show an excellent performance for all cases of practical interest. Then we can use the approximations to do some sensitivity analysis of which results we report. The organization of this monograph is as follows.

In chapter 1 we study the basic model in which excess demand is backlogged. We express the service levels under a given (m,M)-rule in terms of so-called basic functions for which approximations are derived. Our main tools are asymptotic results from renewal theory, and results for ladder height distributions in a random walk where the underlying jump distribution has an exponential tail, cf. also appendix A. Numerical results are presented showing the accuracy of the approximations. We investigate the sensitivity of the switching level m to the underlying demand distribution when keeping the difference M-m fixed and assuming a given service level constraint.

In chapter 2 we analyse the other "extreme" production-inventory model in which excess demand is lost. We derive exact relations between the basic functions associated with the lost-sales model, and those associated with the backlog model. Using these exact relations and the approximations given in chapter 1, we can obtain tractable expressions for the service levels in the lost-sales model. Again the accuracy of the approximations is tested. We also make some comments on the sensitivity of the switching level m to the arrival rate λ when keeping the first and second moment of the demand per unit time fixed and assuming a given service level constraint.

In chapter 3 we consider the model studied in Doshi et al [1978], in which excess demand can be backlogged up to a given amount, and demand in excess of this is lost. Or, equivalently, this model assumes that customers whose demands are backlogged are willing to wait only a fixed amount of time, and leave with the amount that has been produced on their behalf during this waiting time. Using the Markov property of the exponential interarrival times we derive exact relations between this model and the models discussed in the chapters 1 and 2. From these relations approximations can be derived for the operating characteristics of the system.

Chapter 4 deals with another instance of customer impatience. Customers arriving at the production facility wait until their demand is satisfied completely, unless the backlog at the time of arrival exceeds some fixed constant. In the latter case they leave immediately. Equivalently, this model assumes that customers whose demands cannot be met directly from stock on hand leave the system after a fixed amount of time if by that time the production facility has not yet started to produce on their behalf. Hence in this model demand is either completely satisfied or completely lost. For this model we arrive at tractable expressions for the service levels by deriving relations between the basic functions associated with this model and those associated with the models discussed in the chapters 1 and 3.

In chapter 5 we derive accurate and tractable approximations to the average costs per unit time under the cost structure consisting of linear holding costs and fixed switching costs. Using these results we calculate an approximately average-cost optimal (m,M)-rule within the class of (m,M)-rules satisfying a given service level constraint. Next, we numerically verify that the average costs of the (m,M)-rule with M-m equal to the economic production quantity are within 5% of the optimal average costs. The numerical results reveal that the optimal difference M-m becomes insensitive to the service level constraint when the required service level gets high. Using an asymptotic estimate of the average costs under a given (m,M)-rule we derive an approximate expression for the optimal difference M-m. This approximately optimal value of M-m, say $\tilde{\Delta}^*$, is independent of the service measure considered and, moreover, independent of the way excess demand is handled. Further, the average costs of the (m,M)-rule having $M-m=\tilde{\Delta}^*$ and satisfying the service level constraint are within 1% of the optimal average costs. Thus our results show that the sequential determination of M-m and m leads to satisfactory results that are usually close to the optimal results.

In the models considered in the chapters 1 to 4 it is assumed that the time to switch from one production rate to another is negligible. In chapter 6 we study the backlog and lost-sales models with intermittent production (i.e. the slow production rate is zero), where it takes a positive setup time to turn the production on. Using the approximations derived in the chapters 1 and 2 for the backlog and lost-sales model, and applying a two-moment approximation for the distribution of the demand in the setup time, we end up with tractable expressions for the service levels.

Expressions for the average holding and switching costs are given. Numerical results are presented to indicate the accuracy of the approximations, the sensitivity of the switching level m to the underlying demand size distribution, and the sensitivity of the approximately average-cost optimal (m,M)-rule to the setup time.

The production-inventory models discussed in the chapters 1 to 6 assume an infinite storage capacity. Using results from chapter 7 an analogous discussion can be given for finite storage capacity production-inventory models. In chapter 7 we consider a different but related inventory control model. A dam model is discussed in which the content is released at one out of two possible release rates. Inputs occur at epochs generated by a Poisson process. The input sizes have a general probability distribution function. Expressions are derived for service levels of service measures such as the fraction of input that is lost by overflows and the fraction of time that the dam is empty, as well as for the average content of the dam. Both the infinite and the finite capacity dam model are dealt with. The approximations are validated by computer simulation.

In this monograph we restrict our attention to production-inventory models in which the inventory is controlled by an (m,M)-rule. A rule of this simple form is easy to implement in practical situations. A question that remains is whether such a simple rule is optimal among the class of all possible control rules. To our knowledge this problem is in its generality still open, although some results have been obtained in Doshi [1978] and Vickson [1982].

To conclude this introduction we hope that our analysis of the basic models for production-inventory problems with continuous production may provide helpful tools for further research on challenging problems such as production-inventory problems with perishable goods or with multiple products.

. ANALYSIS OF THE BASIC MODEL WITH BACKORDERING OF EXCESS DEMAND.

In this chapter we will consider the single-product production-inventory model where excess demand is backlogged. We develop the basic tools to attack the problem of finding computationally tractable results for this model and its extensions. These extensions are further analyzed in subsequent chapters.

We focus on service measures. As we derive expressions for these service measures we make use of the fact that the inventory process under consideration is regenerative. Throughout this monograph we freely quote standard results from the theory of regenerative processes; see Çinlar [1971], Ross [1970] and Stidham [1972]. Also, to find computationally tractable results we will exploit asymptotic methods for random walks and renewal processes; see Feller [1971]. These methods are summarized in Appendix A.

1.1. The model and preliminaries.

The single-product production-inventory problem to be considered is characterized by a compound Poisson demand process and two possible production rates. The production is continually added to inventory. Customers arrive according to a Poisson process with rate λ and their demands for the single product are independent random variables having a common probability distribution function F with $F(0)=0$. Let the generic random variable D denote the size of a single demand, i.e.

$$P\{D \leq x\} = F(x), \qquad x \geq 0.$$

The demands are independent of the arrival process. We assume that excess demand is backlogged.

At any point in time items are continually added to the inventory at one out of two possible rates π_1 and π_2 with $\pi_1 < \pi_2$. The rates π_1 and π_2 must be interpreted as the differences of a low, respectively high production rate and a constant, possibly zero, demand rate. If this constant demand rate is positive, then the demand process is the sum of a deterministic process and a compound Poisson process. In the sequel π_1 and π_2 will be called production rates. The inventory level decreases with jumps at the arrival epochs of customers and between arrival epochs it increases

or decreases linearly with a slope depending on the production rate

As stated in the introduction we only consider control policies of the following simple structure:

. The production rate is switched from π_1 to π_2 as soon as the inventory level becomes smaller than a critical value $m \geq 0$.

2. The production rate is switched from π_2 to π_1 only when the inventory reaches the critical value $M \geq m$.

Such a control rule will be referred to as an (m,M)-rule. It is assumed that it takes no time to switch from one production rate to the other. For the case of a positive switch time we refer to Chapter 6. We note that for the case of $\pi_1 \geq 0$ the production rate is switched from π_1 to π_2 at arrival epochs only. For the case of $\pi_1 < 0$ the production rate may also be switched from π_1 to π_2 between arrival epochs if the inventory level decreases linearly to the level m.

We assume that the system has an infinite storage capacity. Results for the finite capacity case can be deduced from the results in Chapter 7 where a related dam model is studied. Note that for the case of $\pi_1 \leq 0$ the inventory level cannot exceed the value M.

Our object is to study the long-run behaviour of the inventory process. This requires that the system is "stable". On the one hand the production facility should be able to keep pace with demand sufficiently to prevent exessive shortages, on the other hand inventory should not pile up too much causing high holding costs. Therefore we impose

CONDITION 1.1.1.

$$\pi_1 < \lambda E[D] < \pi_2.$$

Condition 1.1.1 ensures that under the (m,M)-control rule the inventory process cannot drift to ∞ or $-\infty$. This can be proved using the results on random walk theory in Feller [1971], p. 395. Finally we assume for the case of $\pi_1 = 0$ that the probability distribution function F of the random variable D is non-arithmetic (i.e. F is not concentrated on a set $\{0,d,2d,\ldots\}$ for some $d>0$).

The restriction to (m,M)-rules may be motivated as follows: Firstly, from a practical point of view, these (m,M)-rules are easy to implement.

Secondly, though no proof of optimality of these (m,M)-rules with respect to some cost structure exists for the present model, such proofs do exist if the inventory process is a diffusion process; see Doshi [1978] and Vickson [1982] amongst others.

An exact analysis of the general production-inventory model was given in Doshi et al [1978], cf. also Graves and Keilson [1981]. In these papers the criterion was to evaluate the long-run average cost per unit time for a cost structure consisting of fixed setup costs, linear holding costs for inventory, and linear penalty costs for shortages. However, this exact analysis leads to tractable results only for the special case of exponentially distributed demand sizes. In order to obtain practically useful results we should seek a compromise between mathematical and practice-oriented approaches.

Unlike the above studies dealing only with the minimization of costs, we focus on commonly used service measures like the fraction of demand to be met directly from stock on hand. This is motivated by the fact that in practice it is often hard to specify costs associated with shortages. By a probabilistic analysis of the behaviour of the inventory process under a given (m,M)-rule we obtain tractable expressions for a number of service measures of interest. In addition this analysis yields expressions for the average number of switchings of production rate per unit time and for the average on-hand inventory (see chapter 5). These results will enable us to determine the switching levels for the general problem of the minimization of the average switching and holding costs subject to some constraint on the customer service. In particular, when M-m is given, we can determine the switching level m in order to satisfy some service level constraint. It will appear from our results that a sequential determination of M-m and m gives nearly optimal results in practical applications. Here M-m is first determined on the basis of cost considerations only and next the level m is determined on the basis of the service level constraint.

This chapter is further organized as follows. In section 1.2 we give a general outline of the way we use results from the theory of regenerative processes to derive expressions for the service measures. In section 1.3 and 1.4 we derive approximations for the various basic quantities involved by using asymptotic results from renewal theory and random walk theory. In section 1.5 we present the numerical validation of our results. Also we test the sensitivity of the switching level m to the underlying demand distribution for a given value of M-m and a given service level constraint.

1.2. The service measures.

In this section we use the theory of regenerative processes to derive general relations for a number of widely used service measures. Fix an (m,M)-rule with $0 \leq m \leq M$. We shall analyse the inventory process under the given policy. Define for any $t \geq 0$,

$N(t)$:= the number of customers that arrive in $(0,t]$.
$V(t)$:= the total demand in $(0,t]$.
$X(t)$:= the inventory level at time t.
$B(t)$:= the amount of demand in $(0,t]$ that cannot be met directly from stock on hand.
$Q(t)$:= the number of stockouts that occur in $(0,t]$.
$S(t)$:= the number of customers arriving in $(0,t]$ whose total demands cannot be met directly from stock on hand.
$J(t)$:= the amount of time in $(0,t]$ that the inventory is negative.

$$C(t) := -\int_0^t X(s) 1_{\{X(s)<0\}} ds.$$

Here $1_{\{X(s)<0\}}$ is the usual notation for a random variable whose value is 1 if $X(s)<0$ and is 0 otherwise. We say that a stockout occurs if the inventory level drops from a positive value to a non-positive value. The definition of $C(t)$ can be clarified as follows. Imagine that a penalty cost at rate x is incurred when a shortage of x exists. Then $C(t)$ equals the total penalty cost incurred up to time t. In what follows $C(t)$ will be referred to as *the cumulative backlog at time t*.

The above defined stochastic processes underly the following service measures.

(i) α-service measure.

the long-run average number of stockouts per unit time,

$$\lim_{t\to\infty} \frac{Q(t)}{t}.$$

(ii) β-service measure.

the long-run fraction of demand that cannot be met directly from stock on hand,

$$\lim_{t\to\infty} \frac{B(t)}{V(t)} .$$

(iii) γ-service measure.

the long-run fraction of customers whose demands cannot be met directly from stock on hand,

$$\lim_{t\to\infty} \frac{S(t)}{N(t)} .$$

(iv) δ-service measure.

the long-run average backlog at an arbitrary point in time,

$$\lim_{t\to\infty} \frac{C(t)}{t} .$$

By an application of well-known ergodic results from the theory of regenerative processes we show that each of the above limits exists with probability 1 (w.p. 1).

We define

a cycle := the time elapsed between two consecutive epochs at which the inventory level reaches M and the production rate is switched from π_2 to π_1.

Unless stated otherwise we assume that at epoch 0 such a cycle starts. Define for a given (m,M)-rule

T := the next epoch at which the production rate is switched from π_2 to π_1.

Also, let

N := N(T), V := V(T), B := B(T), Q := Q(T), S := S(T),
J := J(T), C := C(T).

Condition 1.1.1 ensures that these random variables have finite expectations. It is easily seen that due to our assumptions on the demand process the *inventory process is regenerative* and hence all the other processes defined above are regenerative. The cycle (0,T] is a regeneration cycle of the inventory process {X(t), t≥0}. Thus we obtain

$$(1.2.1) \qquad \lim_{t\to\infty} \frac{Q(t)}{t} = \frac{E[Q]}{E[T]} \text{ w.p. 1.}$$

$$(1.2.2) \qquad \lim_{t\to\infty} \frac{B(t)}{V(t)} = \frac{E[B]}{E[V]} \text{ w.p. 1.}$$

$$(1.2.3) \qquad \lim_{t\to\infty} \frac{S(t)}{N(t)} = \frac{E[S]}{E[N]} \text{ w.p. 1.}$$

$$(1.2.4) \qquad \lim_{t\to\infty} \frac{C(t)}{t} = \frac{E[C]}{E[T]} \text{ w.p. 1.}$$

$$(1.2.5) \qquad \lim_{t\to\infty} \frac{J(t)}{t} = \frac{E[J]}{E[T]} \text{ w.p. 1.}$$

Equation (1.2.5) gives an expression for the long-run fraction of time the inventory is negative. Also, E[J]/E[T] equals the steady-state probability that the inventory is negative at an arrival epoch. This follows from the property "Poisson arrivals see time averages"; see Wolff [1982].

The service measure (1.2.5) can be related to the β-service measure through

$$(1.2.6) \qquad \frac{E[B]}{E[V]} = \frac{\pi_2}{\lambda E[D]} \cdot \frac{E[J]}{E[T]} .$$

To see this, note that for the compound Poisson demand process the average demand per unit time equals $\lambda E[D]$. On the other hand we have that $\lim_{t\to\infty} V(t)/t = E[V]/E[T]$ w.p. 1. This yields the relation $\lambda E[D] = E[V]/E[T]$. The relation $E[B] = \pi_2 E[J]$ follows by noting that any shortage occurring during a cycle will be gotten quit of at rate π_2 in the same cycle. These relations imply (1.2.6).

We can also express the γ-service measure in terms of the α- and β-service measures. The event that the demand of an arriving customer cannot be met from stock on hand occurs either when he causes a stockout or when the inventory is negative at the time of his arrival. Since E[Q]/E[N] is the long-run fraction of customers, who cause a stockout and E[J]/E[T] is the long-run fraction of customers, who arrive at an epoch at which the inventory is negative, we obtain

$$\frac{E[S]}{E[N]} = \frac{E[Q]}{E[N]} + \frac{E[J]}{E[T]} .$$

Using (1.2.6) and the fact that the property "Poisson arrivals see time averages" implies $E[N] = \lambda E[T]$, we obtain

$$(1.2.7) \qquad \frac{E[S]}{E[N]} = \frac{1}{\lambda} \cdot \frac{E[Q]}{E[T]} + \frac{\lambda E[D]}{\pi_2} \cdot \frac{E[B]}{E[V]} .$$

The relations (1.2.1) - (1.2.7) show that we need tractable expressions for E[T], E[Q], E[B] and E[C]. To evaluate these key elements for the (m,M)-rule we introduce a number of basic functions. We first define the basic functions associated with production rate π_1 and thereafter the basic functions associated with production rate π_2.

We assume that at epoch 0 the inventory level equals x+m, $x \geq 0$, and production rate π_1 is used. We define

$t_1(x)$:= the expected time until the inventory level decreases below m for the first time.

$p(x,u)$:= the probability of having an undershoot greater than u of the level m at the first time the inventory level decreases below m.

$U(x)$:= the undershoot of m at the first time the inventory level decreases below m.

The undershoot of level m is defined as the difference between m and the inventory level immediately after the inventory decreases below m for the first time. It is important to point out that the basic functions $t_1(x)$ and $p(x,u)$ and the random variable $U(x)$ are independent of the switching levels m and M. Note that

$$p(x,u) = P\{U(x)>u\}, \quad u \geq 0.$$

We adopt the convention

$$t_1(0) = 0 \text{ and } U(0) \equiv 0 \text{ for the case of } \pi_1 < 0.$$

Next we define the functions associated with the system while production rate π_2 is used. Assuming that at epoch 0 the inventory level

equals $x \leq M$ and the production rate π_2 is used, we define

$t_2(x)$:= the expected time until the inventory reaches the level M.

$b(x)$:= the expected amount of demand that is backlogged until the inventory reaches the level M (excluding any shortage existing at epoch 0).

$c(x)$:= the expected cumulative backlog at the time at which the inventory reaches the level M.

$q(x)$:= the probability that the inventory level decreases from a positive to a non-positive value before the inventory reaches the level M.

The basic functions $t_2(x)$, $b(x)$, $c(x)$ and $q(x)$ satisfy the boundary conditions $t_2(M) = b(M) = c(M) = q(M) = 0$. Note that as contrasted with $t_1(x)$ and $p(x,u)$ the functions $t_2(x)$, $b(x)$, $c(x)$ and $q(x)$ depend on M. For ease of notation we suppress the dependency of these functions on M.

The following step is to express the quantities E[T], E[B], E[Q] and E[C] in terms of the basic functions. Using the fact that at the beginning of each cycle the inventory equals M and the production rate is switched from π_2 to π_1 and by conditioning on the undershoot of m, we obtain

$$E[T] = t_1(M-m) + \int_0^{\infty} t_2(m-u)\,d_u(1-p(M-m,u)). \tag{1.2.8}$$

$$E[B] = \int_0^{\infty} b(m-u)\,d_u(1-p(M-m,u)) + \int_0^{\infty} (u-m)\,d_u(1-p(M-m,u)). \tag{1.2.9}$$

$$E[C] = \int_0^{\infty} c(m-u)\,d_u(1-p(M-m,u)). \tag{1.2.10}$$

The second term on the right-hand-side of (1.2.9) accounts for the expected shortage occurring when the inventory decreases below m for the first time. To obtain an expression for E[Q] we make the following observations. Since $\pi_2 > \lambda E[D]$ we have that under production rate π_2 the inventory will reach the level 0 with probability 1 for any negative starting value of the inventory. Because of the lack of memory of the exponential interarrival time distribution the past of the system is not relevant when the inventory level reaches the value 0. In other words, the process starts anew each time the inventory level reaches the value 0 and rate π_2 is used. Thus if the current inventory is 0 and π_2 is used then the number of stockouts until the level M

is reached has a geometrical distribution with parameter 1-q(0). Then by conditioning on the undershoot of m we find

$$P\{Q=n\} = \int_0^m q(m-u).q(0)^{n-1}(1-q(0))d_u(1-p(M-m,u))$$

$$+ q(0)^{n-1}(1-q(0))p(M-m,m),$$

implying

$$(1.2.11) \qquad E[Q] = (1-q(0))^{-1}\{p(M-m,m) + \int_0^m q(m-u)d_u(1-p(M-m,u))\}.$$

It remains to find tractable expressions for the basic functions $t_1(x)$, $p(x,u)$, $t_2(x)$, $b(x)$, $c(x)$ and $q(x)$.

1.3. Expressions for p(x,u), E[U(x)] and $t_1(x)$.

In this section we study the inventory process under production rate π_1 and derive tractable expressions for $p(x,u)$, $E[U(x)]$ and $t_1(x)$. Recall that by their definitions these functions do not depend on the particular values m and M of the control rule. The expressions for $E[U(x)]$, $p(x,u)$ and $t_1(x)$ will be obtained by applying the theory of hitting probabilities for random walks and by using asymptotic results from renewal theory. These results from Feller [1971] are summarized in Appendix A. We first define a random walk associated with the inventory process when *always* rate π_1 is used and state some general results for this random walk. Next we separately analyze the cases $\pi_1 \geq 0$ and $\pi_1 < 0$ to obtain approximations for $p(x,u)$, $E[U(x)]$ and $t_1(x)$.

Let us define

τ_1 := the epoch at which the first customer arrives.

τ_n := the time that elapses between the arrival of the (n-1)-th and n-th customer, $n \geq 2$.

D_n := the demand of the n-th arriving customer, $n \geq 1$.

Note that $\{\tau_n\}_{n=1}^{\infty}$ and $\{D_n\}_{n=1}^{\infty}$ are independent sequences of independent identically distributed random variables with

$$P\{\tau_n \leq t\} = 1-e^{-\lambda t}, \quad t \geq 0,\ n \geq 1,$$

$$P\{D_n \leq x\} = F(x), \quad x \geq 0,\ n \geq 1.$$

Next we define the following random walk $\{S_n\}$ on $\mathbb{R}$,

$$S_0 := 0, \quad S_n := \sum_{i=1}^{n} X_i, \quad n \geq 1,$$

where

$$X_n := D_n - \pi_1 \tau_n, \quad n \geq 1.$$

Note that since rate π_1 is always used X_n is the net decrease of the inventory between the arrival of the (n-1)-th and n-th customer. Also, define the sequence of ladder points (ζ_k, Z_k) by

$$\zeta_0 := 0, \quad \zeta_k := \min\{n \mid n>\zeta_{k-1}, S_n > S_{\zeta_{k-1}}\}, \quad k \geq 1$$

and

$$Z_k := S_{\zeta_k}, \quad k \geq 0.$$

We can interprete ζ_k as the k-th arrival epoch at which the inventory level falls below the smallest value attained so far, while $X(0)-Z_k$ is the new smallest value of the inventory level attained at epoch ζ_k. These interpretations are only valid because of the fact that production rate π_1 is always used.

Since $\pi_1 < \lambda E[D]$ we have that $0<E[X_1]<\infty$, implying that the sequences $\{\zeta_k\}$ and $\{Z_k\}$ are renewal processes having a proper probability distribution for the "interarrival times" $\zeta_k-\zeta_{k-1}$ and Z_k-Z_{k-1}, $k\geq 1$. Also (cf. Appendix A), $E[\zeta_1]$ and $E[Z_1]$ are finite and

$$(1.3.1) \qquad E[Z_1] = E[\zeta_1]\, E[X_1] = E[\zeta_1](E[D]-\pi_1/\lambda).$$

Define

$$N(x) := \min\{n \mid S_n > x\}, \quad x \geq 0,$$

$$N^*(x) := \min\{k \mid Z_k > x\}, \quad x \geq 0.$$

It follows from these definitions that

$$(1.3.2) \qquad N(x) = \sum_{i=1}^{N^*(x)} (\zeta_i - \zeta_{i-1}), \quad x \geq 0.$$

Since $N^*(x)$ is a stopping time for the sequence $\{\zeta_k\}$, an application of Wald's equation yields

$$(1.3.3) \qquad E[N(x)] = E[N^*(x)]\, E[\zeta_1], \quad x \geq 0.$$

Note that $M^*(x)$, defined by

$$M^*(x) := E[N^*(x)],$$

is the renewal function associated with the renewal process $\{Z_k\}$. Now a well-known result from renewal theory states that

$$(1.3.4) \qquad \lim_{x\to\infty} P\{Z_{N^*(x)} - x > u\} = \frac{1}{E[Z_1]} \int_u^\infty P\{Z_1 > y\}dy.$$

We can interprete $Z_{N^*(x)} - x$ as the "residual life at epoch x" for the renewal process $\{Z_k\}$.

The above results enable us to analyse the inventory process assuming that production rate π_1 is used. Before applying these results we first give a general relation between $t_1(x)$ and $E[U(x)]$. Towards this end we define for $x \geq 0$

$T_1(x)$ = the time until the inventory level decreases below m, when the initial inventory is x+m.

Then it is immediately clear that

$$(1.3.5) \qquad t_1(x) = E[T_1(x)].$$

Since at epoch $T_1(x)$ the inventory level is undershot by an amount $U(x)$ we have

(1.3.6) $\quad X(T_1(x)) = m-U(x).$

On the other hand we note that the net decrease of the inventory level in $(0,T_1(x)]$ equals the total amount of demand in $(0,T_1(x)]$ minus the total production in $(0,T_1(x)]$,

(1.3.7) $\quad X(0) - X(T_1(x)) = V(T_1(x)) - \pi_1 T_1(x).$

Using $X(0) = x+m$ equations (1.3.5)-(1.3.7) together imply

(1.3.8) $\quad x+m - (m-E[U(x)]) = E[V(T_1(x))] - \pi_1 t_1(x).$

Next we use the property "Poisson arrivals see time averages" to obtain

(1.3.9) $\quad E[V(T_1(x))] = \lambda E[D] t_1(x).$

By combining (1.3.8) and (1.3.9) we obtain

(1.3.10) $\quad t_1(x) = \frac{x+E[U(x)]}{\lambda E[D]-\pi_1}.$

Noting that

(1.3.11) $\quad E[U(x)] = \int_0^\infty p(x,u)du$

it follows that *it suffices to find a tractable expression for p(x,u)*. We distinguish between the case of $\pi_1 \geq 0$ and the case of $\pi_1 < 0$.

Case 1. $\pi_1 \geq 0$:

We first note that for the case of $\pi_1 \geq 0$ the level m can be downcrossed only at arrival epochs. We further observe that the inventory level immediately after the n-th arrival equals $x+m-S_n$, $1 \leq n \leq N(x)$. Using this and the definition of p(x,u) we find

$$p(x,u) = P\{S_{N(x)}-x>u\}, \quad x \geq 0, \; u \geq 0.$$

Clearly we have that $S_{N(x)}-x = Z_{N^*(x)}-x$ and hence

(1.3.12) $\quad p(x,u) = P\{Z_{N^*(x)}-x>u\}, \quad x \geq 0,\ u \geq 0.$

Equation (1.3.12) implies that p(x,u) is completely determined by the renewal process $\{Z_k\}$. We first observe that $N^*(0)=1$ with probability 1 and hence

(1.3.13) $\quad p(0,u) = P\{Z_1>u\}.$

By applying the asymptotic result (1.3.4) we obtain

(1.3.14) $\quad \lim_{x\to\infty} p(x,u) = \frac{1}{E[Z_1]} \int_u^\infty P\{Z_1>y\}dy.$

Equations (1.3.13) and (1.3.14) hold because of the lack of memory of the Poisson arrival process: It is not relevant how long the current interarrival time at epoch 0 is already in progress. The results (1.3.13) and (1.3.14) would also hold for an arbitrary interarrival time distribution, when it is assumed that an arrival occurs at epoch 0 and X(0)=x+m.

We finally wish to express (1.3.13) and (1.3.14) in terms of the demand size distribution function F. At this point we have to distinguish between the cases of $\pi_1=0$ and $\pi_1>0$.

Case 1 (i). $\pi_1=0$:

Then the random walk $\{S_n\}$ is a renewal process and $Z_k=S_k$, $k\geq 0$. Hence it follows that

$$P\{Z_1>u\} = 1-F(u), \quad u \geq 0.$$

Consequently, it follows from (1.3.11), (1.3.13) and (1.3.14) that

(1.3.15) $\quad p(0,u) = 1-F(u), \quad u \geq 0,$

(1.3.16) $\quad E[U(0)] = E[D],$

and

$$\lim_{x\to\infty} p(x,u) = \frac{1}{E[D]} \int_u^\infty (1-F(y))dy, \quad u \geq 0, \tag{1.3.17}$$

$$\lim_{x\to\infty} E[U(x)] = E[D^2](2E[D])^{-1}. \tag{1.3.18}$$

Case 1 (ii). $\pi_1 > 0$:

In this case $\{S_n\}$ is not a renewal process. It is immediately verified from the definition of X_1 that the distribution of X_1 has the special form,

$$P\{X_1 \leq x\} = \begin{cases} e^{\frac{\lambda}{\pi_1}x} \int_0^\infty e^{-\frac{\lambda}{\pi_1}y} dF(y), & x < 0 \\ \int_x^\infty F(y) \frac{\lambda}{\pi_1} e^{-\frac{\lambda}{\pi_1}(y-x)} dy, & x \geq 0 \end{cases} \tag{1.3.19}$$

We observe that $P\{X_1 \leq x\}$ has an exponential left tail. This enables us to apply the results in Feller [1971], p. 405 (cf. Appendix A), for this special type of random walk. We have

$$P\{Z_1 > x\} = \frac{\lambda}{\pi_1} \int_0^\infty e^{-s^* y}(1-F(y+x))dy. \tag{1.3.20}$$

Here s^* is defined as the unique positive root of

$$s - \frac{\lambda}{\pi_1}(1 - \int_0^\infty e^{-sy} dF(y)) = 0. \tag{1.3.21}$$

From (1.3.20) we find

$$E[Z_1] = (\lambda E[D] - \pi_1)/(\pi_1 s^*), \tag{1.3.22}$$

$$E[Z_1^2] = 2\lambda/\pi_1 [E[D^2]/(2s^*) - E[D]/s^{*2} + \pi_1/(\lambda s^{*2})]. \tag{1.3.23}$$

Substituting (1.3.20), (1.3.22) and (1.3.23) into (1.3.11), (1.3.13) and (1.3.14) we obtain

$$p(0,u) = \frac{\lambda}{\pi_1} \int_0^\infty e^{-s^* y}(1-F(y+u))dy, \quad u \geq 0, \tag{1.3.24}$$

$$E[U(0)] = \frac{\lambda E[D] - \pi_1}{\pi_1 s^*} \tag{1.3.25}$$

and

$$(1.3.26)\qquad \lim_{x\to\infty} p(x,u) = \frac{\lambda}{\lambda E[D]-\pi_1}\int_0^\infty (1-e^{-s^* y})(1-F(y+u))dy, \quad u \geq 0,$$

$$(1.3.27)\qquad \lim_{x\to\infty} E[U(x)] = \frac{\lambda E[D^2]}{\lambda E[D]-\pi_1} - \frac{1}{s^*}.$$

We now turn our attention to the case of $\pi_1<0$.

Case 2. $\pi_1<0$:

In this case the equalities (1.3.13) and (1.3.14) do not hold, since the level m can be downcrossed between two arrival epochs. Therefore we take a closer look at the possible events in the interval $(\sum_{n=1}^{N(x)-1}\tau_n, \sum_{n=1}^{N(x)}\tau_n]$, being the interval in which the level m is downcrossed.

We first note that, as in the case of $\pi_1=0$, the random walk $\{S_n\}$ is again a renewal process with $Z_k=S_k$, $k\geq 0$. As stated in section 1.2 we have $U(0)\equiv 0$, implying

$$(1.3.28)\qquad p(0,u) = 0 \quad \text{for all } u \geq 0.$$

Hence to avoid trivialities we now assume that $X(0)=x+m$ with $x>0$. It easily follows that if $x-S_{N(x)-1} < -\pi_1\tau_{N(x)}$ then the level m is downcrossed in the interval $(\sum_{n=1}^{N(x)-1}\tau_n, \sum_{n=1}^{N(x)}\tau_n)$ and the undershoot U(x) equals 0. Thus, for x>0,

$$(1.3.29)\qquad U(x) = \begin{cases} 0 & x-S_{N(x)-1} \leq -\pi_1\tau_{N(x)} \\ S_{N(x)}-x & x-S_{N(x)-1} > -\pi_1\tau_{N(x)} \end{cases}$$

We adopt the convention that U(x)=0 if the null-event $\{x-S_{N(x)-1} = -\pi_1\tau_{N(x)}\}$ occurs. Applying the same arguments as in Feller [1971], p. 369, we find

$$P\{U(x)>u\} = \int_0^x \int_0^{(x-y)/(-\pi_1)} [1-F(x+u-(y-\pi_1 z))]\lambda e^{-\lambda z} dz dM^*(y), \quad u \geq 0,$$

where we used $E[N(x)] = E[N^*(x)] = M^*(x)$. After a change of variable we obtain

$$P\{U(x)>u\} = \int_0^x H(x-y)dM^*(y), \quad u \geq 0,\ x > 0,$$

with H(y) defined by

$$H(y) := \int_0^y (1-F(y+u-w)\frac{\lambda}{\pi_1}e^{-\frac{\lambda}{\pi_1}w}dw.$$

By applying the Key Renewal Theorem (cf. Feller [1971], p. 363) and using (1.3.29) we find the asymptotic result

$$\lim_{x\to\infty} p(x,u) = \frac{\lambda}{\lambda E[D]-\pi_1}\int_u^\infty (1-F(y))dy, \quad u \geq 0. \tag{1.3.30}$$

Letting $u\to 0$ in (1.3.30) we arrive at

$$\lim_{x\to\infty} P\{U(x)=0\} = \frac{-\pi_1}{\lambda E[D]-\pi_1}. \tag{1.3.31}$$

It follows from (1.3.30) that

$$\lim_{x\to\infty} E[U(x)] = \frac{\lambda E[D^2]}{2(\lambda E[D]-\pi_1)}. \tag{1.3.32}$$

Concluding, for each of the three cases $\pi_1>0$, $\pi_1<0$ and $\pi_1=0$ we have found exact expressions for $p(0,u)$ and $E[U(0)]$ and asymptotic expansions for $p(x,u)$ and $E[U(x)]$ as $x\to\infty$. It is a matter of some algebra to verify that for $\pi_1\to 0$ the results for either of the cases $\pi_1>0$ and $\pi_1<0$ coincide indeed with those for the case of $\pi_1=0$. To do so use that $\pi_1 s^* \to \lambda$ and $s^* \to \infty$ as $\pi_1 \downarrow 0$ and rewrite (1.3.24) as

$$p(0,u) = \frac{\lambda}{\pi_1 s^*}\int_u^\infty (1-e^{-s^*(y-u)})dF(y), \quad u \geq 0.$$

We now turn to the specifications of the above results for a given (m,M)-rule. First we define the random variable U,

$$U := U(M-m).$$

Now we distinguish between the cases m=M and m<M.

Case M=m:

In this case we can apply the exact results derived for $p(0,u)$ and $E[U(0)]$. It follows from (1.3.10), (1.3.15), (1.3.16), (1.3.24), (1.3.25), (1.3.29) and $U(0)\equiv 0$ for $\pi_1<0$ that we can summarize the following results.

$$(1.3.33)\qquad t_1(0) = \begin{cases} 0 & \text{when } \pi_1 < 0 \\ 1/\lambda & \text{when } \pi_1 = 0 \\ 1/(\pi_1 s^*) & \text{when } \pi_1 > 0 \end{cases}$$

For any $u \geq 0$,

$$(1.3.34)\qquad p(0,u) = \begin{cases} 0 & \text{when } \pi_1 < 0 \\ 1-F(u) & \text{when } \pi_1 = 0 \\ (\lambda/\pi_1)\int_0^\infty e^{-s^* y}(1-F(y+u))dy & \text{when } \pi_1 > 0 \end{cases}$$

$$(1.3.35)\qquad E[U] = \begin{cases} 0 & \text{when } \pi_1 < 0 \\ E[D] & \text{when } \pi_1 = 0 \\ (\lambda E[D]-\pi_1)/(\pi_1 s^*) & \text{when } \pi_1 > 0 \end{cases}.$$

Case M>m:

In this case we need expressions for $t_1(M-m)$ and $p(M-m,u)$. In general we can only give approximate expressions for these quantities. These approximations are based on the asymptotic expressions for $p(x,u)$ and $E[U(x)]$ as $x\to\infty$.

Let us consider the case of $\pi_1\geq 0$. The asymptotic expressions for $p(x,u)$ and $E[U(x)]$ have been derived from equation (1.3.4). Now we address ourselves to the following problem. Given a renewal process $\{Z_k\}$ determine empirically an estimate for that value of x_0 yielding a good approximation

$$P\{Z_{N^*(x_0)}-x_0>u\} \cong \frac{1}{E[Z_1]}\int_u^\infty P\{Z_1>y\}dy$$

with $N^*(x)=\min\{k \mid Z_k>x\}$. To estimate x_0 we used as criterion when the first two moments of the above two distributions were sufficiently close to one another, where we estimated the first two moments of the distribution on the right-hand side of the above relation by computer simulation. From our extensive numerical investigations we found that an appropriate choice for x_0 is given by

$$(1.3.36) \qquad x_0 = \begin{cases} E[Z_1] & \text{when } c_{Z_1}^2 \leq 1 \\ 1\tfrac{1}{2}\, c_{Z_1}^2 E[Z_1] & \text{when } c_{Z_1}^2 > 1 \end{cases}$$

where $c_{Z_1}^2$ denotes the squared coefficient of variation of Z_1.

We noted before that the sequence $\{Z_k\}$ of ladder heights associated with the random walk $\{S_n\}$ constitutes a renewal process. Then it follows from (1.3.14) and the above discussion that for the case of $\pi_1 \geq 0$ $p(M-m,u) \cong (E[Z_1])^{-1} \int_u^\infty (1-P\{Z_1>y\})dy$ if $M-m \geq x_0$ with x_0 given by (1.3.36)

For the case of $\pi_1 < 0$ we cannot apply the results of the above discussion, since, as opposed to the case of $\pi_1 \geq 0$, the level m can be downcrossed between arrival epochs. However, numerical results indicated that the first two moments of $p(x_0,u)$ are reasonably well approximated by the first two moments of the asymptotic undershoot distribution given by the right-hand side of (1.3.30) when

$$x_0 = \begin{cases} E[D]-\pi_1/\lambda & \text{when } c_D^2 \leq 1 \\ 1\tfrac{1}{2}\, c_D^2(E[D]-\pi_1/\lambda) & \text{when } c_D^2 > 1 \end{cases}$$

Hence for the case of $M-m>0$ we restrict ourselves to (m,M)-policies satisfying

Condition 1.3.1.

For the case of $\pi_1 \leq 0$,

$$M-m \geq \begin{cases} E[D]-\pi_1/\lambda & \text{when } c_D^2 \leq 1 \\ 1\tfrac{1}{2}\, c_D^2(E[D]-\pi_1/\lambda) & \text{when } c_D^2 > 1 \end{cases}$$

and for the case of $\pi_1 > 0$,

$$M-m \geq \begin{cases} E[Z_1] & \text{when } c_{Z_1}^2 \leq 1 \\ 1\tfrac{1}{2}\, c_{Z_1}^2 E[Z_1] & \text{when } c_{Z_1}^2 > 1 \end{cases}$$

where $c_{Z_1}^2 = (E[Z_1^2]-(E[Z_1])^2)/(E[Z_1])^2$ and $E[Z_1]$ and $E[Z_1^2]$ are given by by (1.3.22) and (1.3.23) respectively. This restriction is reasonable for applications where switching costs are involved.

Then, using (1.3.10), (1.3.17), (1.3.18), (1.3.26), (1.3.27), (1.3.30), (1.3.31) and (1.3.32), we find the following approximations,

Approximation 1.3.1.

$$t_1(M-m) \cong \begin{cases} \dfrac{M-m}{\lambda E[D]-\pi_1} + \dfrac{\lambda E[D^2]}{2(\lambda E[D]-\pi_1)^2} & \text{when } \pi_1 \leq 0 \\ \dfrac{M-m}{\lambda E[D]-\pi_1} + \dfrac{\lambda E[D^2]}{2(\lambda E[D]-\pi_1)^2} - \dfrac{1}{s^*(\lambda E[D]-\pi_1)} & \text{when } \pi_1 > 0 \end{cases}$$

Approximation 1.3.2.

$$p(M-m,u) \cong \begin{cases} \lambda/(\lambda E[D]-\pi_1) \int_u^\infty (1-F(y))dy & \text{when } \pi_1 \leq 0 \\ \lambda/(\lambda E[D]-\pi_1) \int_u^\infty (1-F(y))(1-e^{-s^*(y-u)}dy & \text{when } \pi_1 > 0 \end{cases}$$

Approximation 1.3.3.

$$E[U] \cong \begin{cases} \dfrac{\lambda E[D^2]}{2(\lambda E[D]-\pi_1)} & \text{when } \pi_1 \leq 0 \\ \dfrac{\lambda E[D^2]}{2(\lambda E[D]-\pi_1)} - \dfrac{1}{s^*} & \text{when } \pi_1 > 0 \end{cases}$$

The constant s^* is determined by (1.3.21).

In table 1.3.1 we give for $E[U]$ and c_U^2 the approximate values and the actual values obtained by computer simulation. Here c_U^2 is the squared coefficient of variation of U,

$$c_U^2 := \frac{E[U^2]-(E[U])^2}{(E[U])^2} .$$

$E[U^2]$ is computed from approximation 1.3.2. The value of M-m is set equal to the lower bound in condition 1.3.1. We considered the following demand distributions:

(i) deterministic demand ($c_D^2 = 0$).

(ii) Erlang-2 demand ($c_D^2 = 0.5$).

(iii) hyperexponential demand with balanced means ($c_D^2 = 2$), i.e.

$$F(x) = 1 - pe^{-\mu_1 x} - (1-p)e^{-\mu_2 x}, \quad x > 0, \ \frac{p}{\mu_1} = \frac{1-p}{\mu_2}.$$

Here c_D^2 denotes the squared coefficient of variation of the demand D. In all examples we have chosen $\lambda=1$, $E[D]=1$ and π_1 has the four values -2, -0.5, 0, 0.5.

Table 1.3.1. Accuracy of approximations for $E[U]$ and c_U.

	$c_D^2 = 0$		$c_D^2 = 0.5$		$c_D^2 = 2$	
π_1	$E[U]_{ap}$	$E[U]_{act}$	$E[U]_{ap}$	$E[U]_{act}$	$E[U]_{ap}$	$E[U]_{act}$
-2	0.17	0.17	0.25	0.25	0.50	0.51
-0.5	0.33	0.32	0.50	0.51	1.00	0.98
0			0.75	0.76	1.50	1.43
0.5	0.37	0.32	0.69	0.70	1.78	1.77
π_1	$c_{U,ap}$	$c_{U,act}$	$c_{U,ap}$	$c_{U,act}$	$c_{U,ap}$	$c_{U,act}$
-2	1.73	1.73	2.08	2.07	2.65	2.65
-0.5	1.00	1.12	1.29	1.28	1.73	1.75
0			0.88	0.88	1.29	1.32
0.5	0.66	0.70	0.92	0.91	1.19	1.22

In table 1.3.1 we omitted the values of $E[U]$ and c_U for the case of $\pi_1=0$ and $c_D^2=0$, since for this case it is trivial to compute the exact undershoot distribution.

Remark 1.3.1.

Consider the special case of exponentially distributed demand size with mean $1/\mu$, i.e.

$$F(x) = 1-e^{-\mu x}, \quad x \geq 0.$$

It follows from (1.3.15), (1.3.21) and (1.3.24) that for the case of $\pi_1 \geq 0$

$$p(0,u) = P\{Z_1 > u\} = e^{-\mu u}, \quad u \geq 0.$$

This implies that the renewal process $\{Z_k\}$ is a Poisson process with rate μ. So by the lack of memory of the Poisson process,

$$p(x,u) = P\{Z_{N^*(x)} - x > u\} = e^{-\mu u}, \quad u \geq 0, \; x \geq 0.$$

This result together with (1.3.10) and (1.3.11) implies

$$t_1(x) = \frac{\mu x + 1}{\lambda - \pi_1 \mu}, \quad x \geq 0.$$

Using the above results it can be seen that for the case of $\pi_1 \geq 0$ the approximations 1.3.1-1.3.3 are exact when the demand has an exponential distribution.

1.4. Expressions for $t_2(x)$, $q(x)$, $b(x)$ and $c(x)$.

In this section we derive approximations for the basic functions associated with the evolution of the process $\{X(t), t \geq 0\}$ during the time that production rate π_2 is used. The Key Renewal Theorem plays again a major part in the analysis. We also use some results from queueing theory.

There is a fundamental difference between the approximations given in section 1.3 and those to be derived in this section. In section 1.3 we could justify condition 1.3.1 since M-m is typically large in practical applications when switching costs are involved. This observation provided solid ground for the application of asymptotic results. However, in this section we have to find approximations for functions associated with the inventory level immediately after the switching level m has been downcrossed. This inventory level can have any value less than m, so that useful approximations have to be found for all starting levels $x \leq m$ when using rate π_2.

Throughout this section we assume that at epoch 0 the inventory level equals $x \leq M$, i.e. $X(0)=x$, and production rate π_2 is used. We first derive an exact expression for $t_2(x)$. Let us define

$T_2(x)$:= the time until the inventory level reaches the value M, $x \leq M$.

Then we have by definition that $t_2(x)=E[T_2(x)]$. Since excess demand is backlogged it can be seen that the process $\{M-X(t),\ 0\leq t\leq T_2(x)\}$ corresponds to the workload process in the following M/G/1-type queueing system. Jobs arrive according to a Poisson process with rate λ and the job sizes are independent random variables with common distribution function F. Work is processed at a constant rate π_2 whenever the system is non-empty. The initial workload is M-x. Then it follows from a well-known result (see e.g. Tijms [1977]) that

$$t_2(x) = \frac{M-x}{\pi_2-\lambda E[D]}, \quad x \leq M. \tag{1.4.1}$$

An alternative derivation of (1.4.1) follows the lines of the derivation of (1.3.10), based on the conservation of flow and the lack of memory of the Poisson arrival process.

Next we focus on the hitting probability q(x). We recall that q(x) depends on M. Let us assume for the moment that $M=\infty$. This is equivalent to assuming that production rate π_2 is *always* used. Let

$q_\infty(x)$:= the probability that the inventory level will ever decrease from a positive to a non-positive value when production rate π_2 is always used, $x \geq 0$,

be the corresponding hitting probability. We will prove the following relation between q(x) and $q_\infty(x)$,

$$q(x) = \frac{q_\infty(x)-q_\infty(M)}{1-q_\infty(M)}, \quad 0 \leq x \leq M. \tag{1.4.2}$$

Towards this end we note that

$$q_\infty(x) = P\{X(t_0)\leq 0 \text{ for some } t_0>0 \mid X(0)=x\}, \quad x \geq 0.$$

If $0\leq X(0)<M$ and if there exists a t_0 such that $X(t_0)\leq 0$ and $X(t)>0$ for $0<t<t_0$ then there are two mutually exclusive possibilities:

(i) $X(t)<M$ for $0 < t < t_0$.

(ii) $X(t_1)=M$ for some $0 < t_1 < t_0$.

This implies that

$$q_\infty(x) = P\{X(t_0)\leq 0 \text{ for some } t_0>0 \text{ and } 0<X(t)<M \text{ for } 0<t<t_0 \mid X(0)=x\}$$

$$+ P\{X(t_0)\leq 0 \text{ for some } t_0>0 \text{ and } X(t)>0 \text{ for } 0<t<t_0 \tag{1.4.3}$$

$$\text{and } X(t_1)=M \text{ for some } 0<t_1<t_0 \mid X(0)=x\}, \quad 0 \leq x < M.$$

By definition the first term on the right-hand side of (1.4.3) equals $q(x)$ corresponding to $M<\infty$. The second term on the right-hand side of (1.4.3) can be rewritten as follows. By conditioning on the event of reaching M before emptiness we find for all $0\leq x<M$

$$P\{X(t_0)\leq 0 \text{ for some } t_0>0 \text{ and } X(t)>0 \text{ for } 0<t\leq t_0$$
$$\text{and } X(t_1)=M \text{ for some } 0<t_1<t_0 \mid X(0)=x\}$$

$$= P\{X(t_0)\leq 0 \text{ for some } t_0>t_1 \text{ and } X(t)>0 \text{ for } t_1<t<t_0 \mid \tag{1.4.4}$$
$$X(t_1)=M \text{ for some } t_1>0 \text{ and } 0<X(t)<M \text{ for } 0<t<t_1 \text{ and } X(0)=x\}$$

$$\times P\{X(t_1)=M \text{ for some } t_1>0 \text{ and } 0<X(t)<M \text{ for } 0<t<t_1 \mid X(0)=x\}.$$

Due to the compound Poisson demand process it follows that the evolution of the inventory process from any time t onward depends on the history of the inventory process up to time t only through X(t). Hence we have for all $0\leq x<M$

$$P\{X(t_0)\leq 0 \text{ for some } t_0>t \text{ and } X(t)>0 \text{ for } t_1<t<t_0 \mid$$
$$X(t_1)=M \text{ for some } t_1>0 \text{ and } 0<X(t)<M \text{ for } 0<t<t_1 \text{ and } X(0)=x\}$$

$$= P\{X(t_0)\leq 0 \text{ for some } t_0>t_1 \text{ and } X(t)>0 \text{ for } t_1<t<t_0 \mid X(t_1)=M\}$$

$$= q_\infty(M).$$

The last equality follows by taking epoch t_1 as the new time origin. The second term on the right-hand side of (1.4.4) is 1-q(x). Substituting these results in the equations (1.4.3) and (1.4.4) yields

$$q_\infty(x) = q(x) + (1-q(x))q_\infty(M), \quad 0 \leq x < M.$$

This equation and q(M)=0 together imply equation (1.4.2).

So it suffices to find an approximation for $q_\infty(x)$ corresponding to the case of $M=\infty$. For this purpose we note that $q_\infty(x)$ is the classical ruin probability which is extensively studied in the literature. From pp. 377-378 in Feller [1971] (cf. also Cohen [1976], p. 79) we have

$$q_\infty(0) = \lambda E[D]/\pi_2 \tag{1.4.5}$$

$$\lim_{x\to\infty} e^{\delta x} q_\infty(x) = \frac{\pi_2 - \lambda E[D]}{\delta \nu \pi_2} \tag{1.4.6}$$

where δ is defined as the unique positive root of

$$1 - \int_0^\infty e^{\delta y} \frac{\lambda}{\pi_2} (1-F(y))dy = 0 \tag{1.4.7}$$

and ν is given by

$$\nu = \int_0^\infty y e^{\delta y} \frac{\lambda}{\pi_2} (1-F(y))dy. \tag{1.4.8}$$

However, the transcendental equation (1.4.7) has not always a positive root. For instance for the lognormal distribution function and distribution functions with regularly varying tails (i.e. $\lim_{t\to\infty}(1-F(tx))/(1-F(t))=x^\rho$, $\rho\in\mathbb{R}$, $x\in\mathbb{R}^+$) the integral on the left-hand side of (1.4.7) diverges for all $\delta>0$. Fortunately, the equation (1.4.7) is solvable when F has an exponentially fast decreasing tail. Therefore we make the following

Distribution Assumption DA

$$1-F(x) = O(e^{-\kappa x}) \quad \text{for some } \kappa > 0 \ (x\to\infty). \tag{1.4.9}$$

In fact, this assumption is necessary and sufficient to ensure that equation (1.4.7) has a unique positive root. Then also $\nu<\infty$. This assumption is satisfied for many distributions of practical interest, including distributions with a finite support and mixtures of Erlangian distributions. Note that any distribution function with support on $[0,\infty)$ can arbitrarily closely be approximated by a finite mixture of Erlangian distributions, cf. Cox [1955] and Schassberger [1972]. We further note that DA implies that $E[D^k]<\infty$ for all $k\geq 1$.

Now the key idea to all approximations of this section is to find a function of a simple form that couples the behaviour of the basic function under consideration near the origin and its asymptotic behaviour near infinity. There are several coupling possibilities, but our choice will be guided by the simple form of an exact solution for $q_\infty(x)$ when the demand has a so-called K_2-distribution to be discussed below.

Equation (1.4.6) provides some information about the behaviour of $q_\infty(x)$ near the origin. But it is possible to obtain more detailed information. As stated above $q_\infty(x)$ corresponds to the classical ruin probability. It follows from the results derived in Feller [1971], pp. 194-198, that the ruin probability is identical to the waiting time probability in a single-server queueing system with service in order of arrival. More precisely,

$$q_\infty(x) = 1 - W(x), \tag{1.4.10}$$

where $W(x)$ denotes the probability that the actual waiting time of an arbitrary customer is less than x/π_2 in the M/G/1-queue with arrival rate λ, service requirement D and processing rate π_2. Using "Poisson arrivals see time averages" it follows that $W(x)$ equals the probability that at an arbitrary point in time the workload is less than x in this M/G/1-queueing system. Hence

$$\int_0^\infty (1-W(y))dy = \text{the expected amount of work at an arbitrary point in time in the M/G/1-queueing system with arrival rate } \lambda, \text{ service requirement D and processing rate } \pi_2. \tag{1.4.11}$$

It is well-known (see e.g. Tijms [1977]) that the average workload in the M/G/1-queue equals $\frac{1}{2}\lambda E[D^2](\pi_2-\lambda E[D])^{-1}$. Hence

$$(1.4.12) \qquad \int_0^\infty q_\infty(y)dy = \frac{\lambda E[D^2]}{2(\pi_2-\lambda E[D])} .$$

Equations (1.4.5) and (1.4.12) describe the behaviour of $q_\infty(x)$ near (at) the origin, while the equation (1.4.6) describes the asymptotic behaviour of $q_\infty(x)$ as $x\to\infty$.

Next we determine the exact solution of $q_\infty(x)$ for the class of K_2-distributions. A probability distribution function F is called a K_2-distribution function if the Laplace-Stieltjes transform

$$\tilde{F}(s) := \int_0^\infty e^{-sx}dF(x), \quad s \geq 0,$$

of F can be written as

$$\tilde{F}(s) = \frac{1+a_0 s}{1+a_1 s+a_2 s^2}$$

for some constants a_0, a_1 and a_2. The class of K_2-distributions contains hyperexponential distributions and mixtures of Erlang-1 and Erlang-2 distributions with the same scale parameters. We now show that $q_\infty(x)$ is of a simple form when F is a K_2-distribution.

Consider the time interval $(0,\Delta x/\pi_2)$ with Δx small. In this time interval a demand occurs with probability $\lambda\Delta x/\pi_2+o(\Delta x)$, since the customers arrive according to a Poisson process with rate λ. Hence, for each x such that x is a continuity point of the demand size distribution F, we have

$$q_\infty(x) = \frac{\lambda\Delta x}{\pi_2}\{\int_0^x q_\infty(x-y)dF(y)+1-F(x)\}+(1-\frac{\lambda\Delta x}{\pi_2})q_\infty(x+\Delta x)+o(\Delta x), \qquad x \geq 0.$$

Dividing both sides of this relation by Δx, rearranging terms, letting $\Delta x\downarrow 0$ and noting that F has at most a countable number of discontinuity points, we find for almost all $x\geq 0$,

$$(1.4.13) \qquad q_\infty'(x) = -\frac{\lambda}{\pi_2}(1-F(x)) + \frac{\lambda}{\pi_2}q_\infty(x) - \frac{\lambda}{\pi_2}\int_0^x q_\infty(x-y)dF(y).$$

We define

$$\tilde{q}_\infty(s) := \int_0^\infty e^{-sy}q_\infty(y)dy, \quad s > 0,$$

so $\tilde{q}_\infty(s)$ is the ordinary Laplace transform of $q_\infty(x)$. Taking the Laplace-transform on both sides of (1.4.13) it readily follows, using $q_\infty(0) = \lambda E[D]/\pi_2$, that

$$(1.4.14) \qquad \tilde{q}_\infty(s) = \frac{\lambda E[D]-\lambda(1-\tilde{F}(s))/s}{\pi_2 s-\lambda(1-\tilde{F}(s))}, \qquad s > 0.$$

In case F is a K_2-distribution function we find

$$\tilde{q}_\infty(s) = \frac{b_1 s+b_0}{s^2+c_1 s+c_0}, \qquad s > 0,$$

for some constants b_0, b_1, c_0, c_1. Except for the case of exponential demand the equation $s^2+c_1 s+c_0=0$ has two positive roots. By a simple inversion we obtain for some α, β and γ

$$(1.4.15) \qquad q_\infty(x) = \alpha e^{-\beta x} + \gamma e^{-\delta x}, \qquad x \geq 0,$$

when F is a K_2-distribution. The constant δ is defined by equation (1.4.7). Also, it must be true that $\delta<\beta$ in view of the asymptotic expansion (1.4.6).

The exact result (1.4.15) for the K_2-distribution function suggests to approximate $q_\infty(x)$ by

$$(1.4.16) \qquad q_\infty(x) \cong \alpha e^{-\beta x} + \frac{(\pi_2-\lambda E[D])}{\pi_2 \delta \nu} e^{-\delta x}, \qquad x \geq 0,$$

where δ and ν are given by (1.4.7) and (1.4.8). The equation (1.4.16) encompasses the asymptotic behaviour of $q_\infty(x)$ given by (1.4.6). Using (1.4.5), (1.4.6) and (1.4.12) we can solve for the unknown constants α and β, yielding

$$(1.4.17) \qquad \alpha = \frac{\lambda E[D]}{\pi_2} - \frac{(\pi_2-\lambda E[D])}{\pi_2 \delta \nu}$$

$$(1.4.18) \qquad \beta = \left(\frac{\lambda E[D]}{\pi_2} - \frac{(\pi_2-\lambda E[D])}{\pi_2 \delta \nu}\right)\left(\frac{\lambda E[D^2]}{2(\pi_2-\lambda E[D])} - \frac{(\pi_2-\lambda E[D])}{\pi_2 \delta^2 \nu}\right)^{-1}$$

The following two conditions are required for the constants α, β, γ and δ.

(i) $\beta > \delta$.

(ii) $\alpha\beta + \gamma\delta \geq 0$.

Condition (i) should hold in view of (1.4.6), while condition (ii) should hold since $q_\infty(x)$ is a non-increasing function. Except for the class of K_2-distributions, we were not able to characterize further distributions for which (i) and (ii) hold. Therefore we numerically tested many practical distributions satisfying DA. For mixtures of Erlangian distributions we always found that (i) was satisfied. Occasionally we found for such mixtures that condition (ii) was violated. But in these cases the right-hand side of (1.4.16) is slightly increasing only close to the origin, where the maximum is very close to $q_\infty(0)=\lambda E[D]/\pi_2$. This phenomenon apparently does not affect the quality of the approximations as may be concluded from the numerical results in section 1.5. Hence we propose

Approximation 1.4.1. For all $0\leq x\leq M$

$$q(x) = \frac{q_\infty(x) - q_\infty(M)}{1-q_\infty(M)}$$

with

$$q_\infty(x) \cong \alpha e^{-\beta x} + \frac{(\pi_2-\lambda E[D])}{\pi_2 \delta \nu} e^{-\delta x}, \quad x \geq 0,$$

where α and β are given by (1.4.17) and (1.4.18).

Next we derive an approximation for the basic function $b(x)$. First consider the case $x\leq 0$. Let us define

$$T_0(x) := \text{the time until the inventory level reaches the value 0, } x \leq 0.$$

By the same arguments as used to derive (1.4.1) we have

$$E[T_0(x)] = \frac{-x}{\pi_2-\lambda E[D]}, \quad x \leq 0.$$

By definition $b(x)=E[B(T_2(x))]$. Writing $B(T_2(x))=B(T_0(x))+[B(T_2(x))-B(T_0(x))]$ we can again use the fact that $\{X(t),\ t\geq T_0(x)\}$ depends on $\{X(t),\ 0\leq t\leq T_0(x)\}$ only through $X(T_0(x))=0$. Then we find, using the definition of $b(x)$,

$$b(x) = E[B(T_0(x))] + b(0), \quad x \leq 0.$$

Now we express $B(T_0(x))$ and $T_0(x)$ in terms of $\{D_n\}$ and $\{\tau_n\}$,

$$B(T_0(x)) = \sum_{n=1}^{N(T_0(x))} D_n,$$

$$T_0(x) = \sum_{n=1}^{N(T_0(x))+1} \tau_n - \Big(\sum_{n=1}^{N(T_0(x))+1} \tau_n - T_0(x) \Big).$$

It easily follows that $N(T_0(x))$ is a stopping time for $\{D_n\}$ and $N(T_0(x))+1$ is a stopping time for $\{\tau_n\}$. It follows from the Markov property of the exponential interarrival times that

$$E\Big[\sum_{n=1}^{N(T_0(x))+1} \tau_n - T_0(x) \Big] = 1/\lambda.$$

Applying Wald's equation twice yields

$$E[B(T_0(x))] = \lambda E[D]\, E[T_0(x)]$$

and hence

$$b(x) = \frac{-\lambda E[D]x}{\pi_2 - E[D]} + b(0), \quad x \leq 0. \tag{1.4.19}$$

We emphasize the fact that the shortage $-x$ at epoch 0 is not included in $b(x)$.

By applying the same arguments as used to derive (1.4.13) and invoking (1.4.19) we obtain for almost all $0<x<M$,

$$b'(x) = \frac{-\lambda}{\pi_2 - \lambda E[D]} \int_x^\infty (y-x)dF(y) - \frac{\lambda}{\pi_2} b(0)(1-F(x)) + \frac{\lambda}{\pi_2} b(x) - \frac{\lambda}{\pi_2} \int_0^x b(x-y)dF(y). \tag{1.4.20}$$

Let us again consider the special case of $M=\infty$. We define

$$b_\infty(x) := E[B(\infty) \mid X(0)=x], \quad x \in \mathbb{R}$$

i.e. $b_\infty(x)$ equals the expected amount of demand that will go short during the interval $(0,\infty)$ when $X(0)=x$ and always rate π_2 is used. Fix $x<M$. Now it obviously holds that

(1.4.21) $E[B(\infty)|X(0)=x] = E[B(T_2(x))|X(0)=x] + E[B(\infty)-B(T_2(x))|X(0)=x].$

By definition we have

(1.4.22) $b(x) = E[B(T_2(x))|X(0)=x].$

We further note that $X(T_2(x)) = M$ with probability 1.
Conditioning on $T_2(x)=t_M$ and noting that the proces $\{X(t),\ t\geq t_M\}$ depends on its history $\{X(t),\ 0\leq t\leq t_M\}$ only through $X(t_M)$ we find

(1.4.23) $E[B(\infty)-B(T_2(x))|X(0)=x,\ T_2(x)=t_M] = E[B(\infty)-B(t_M)|X(t_M)=M]=b_\infty(M).$

Taking the expectation with respect to $T_2(x)$ in (1.4.23) we find

(1.4.24) $E[B(\infty)-B(T_2(x)|X(0)=x] = b_\infty(M).$

By combining the relations (1.4.21), (1.4.22) and (1.4.24) we get a similar result as (1.4.2),

(1.4.25) $b(x) = b_\infty(x) - b_\infty(M), \quad x \leq M.$

Since under production rate π_2 the inventory process has a positive drift to infinity, it is intuitively clear that

(1.4.26) $\lim_{x\to\infty} b_\infty(x) = 0.$

This result can be proved rigorously by using arguments based on results for terminating renewal processes induced by ladder heights. But there is more to say about the asymptotic behaviour of $b_\infty(x)$. From equation (1.4.20) we derive by integration that

$$b_\infty(x) = w(x) + \frac{\lambda}{\pi_2}\int_0^x \{b_\infty(y) - \int_0^y b_\infty(y-z)dF(z)\}dy, \quad x \geq 0,$$

where

$$(1.4.27)\qquad w(x) = b_\infty(0) - \frac{\lambda}{\pi_2-\lambda E[D]} \int_0^x \int_y^\infty (1-F(z))dzdy - \frac{\lambda}{\pi_2} b_\infty(0) \int_0^x (1-F(y))dy,$$

$$x \geq 0.$$

Using

$$\frac{d}{dx}\{\int_0^x f(x-y)(1-F(y))dy\} = f(x) - \int_0^x f(x-y)dF(y)$$

for any continuous function f, we find

$$(1.4.28)\qquad b_\infty(x) = w(x) + \frac{\lambda}{\pi_2} \int_0^x b_\infty(x-y)(1-F(y))dy, \quad x \geq 0.$$

Equation (1.4.28) is a so-called defective renewal equation since $\lambda E[D]/\pi_2 < 1$ and thus (see p. 375 in Feller [1971]),

$$(1.4.29)\qquad \lim_{x\to\infty} b_\infty(x) = \lim_{x\to\infty} \frac{w(x)}{1-\frac{\lambda E[D]}{\pi_2}}.$$

The equations (1.4.26), (1.4.27) and (1.4.29) together imply that

$$(1.4.30)\qquad b_\infty(0) = \frac{\lambda\pi_2 E[D^2]}{2(\pi_2-\lambda E[D])^2}.$$

The asymptotic behaviour of $b_\infty(x)$ as $x\to\infty$ can be obtained by using the Key Renewal Theorem. By the definition of δ and ν we have that

$$G(x) := \int_0^x e^{\delta y} \frac{\lambda}{\pi_2} (1-F(y))dy, \quad x \geq 0,$$

is a proper distribution function, while

$$\nu = \int_0^\infty y dG(y)$$

is the first moment of the distribution function G. By multiplying both sides of (1.4.28) by $e^{\delta x}$ and using the definition of G we obtain the proper renewal equation,

$$e^{\delta x} b_\infty(x) = e^{\delta x} w(x) + \int_0^x e^{\delta(x-y)} b_\infty(x-y)dG(y), \quad x \geq 0.$$

Applying the Key Renewal Theorem yields

$$(1.4.31)\qquad \lim_{x\to\infty} e^{\delta x} b_\infty(x) = \frac{1}{\nu} \int_0^\infty e^{\delta y} w(y)dy.$$

To evaluate the integral in (1.4.31), we first rewrite w(x) in a more convenient form. Writing $b_\infty(0)=(1-\lambda E[D]/\pi_2)b_\infty(0)+\lambda E[D]/\pi_2 b_\infty(0)$ and using (1.4.30) and $E[D^k]=k\int_0^\infty y^{k-1}[1-F(y)]dy$ for k=1,2, we obtain after some straightforward algebra

$$w(x) = \frac{\lambda}{\pi_2-\lambda E[D]} \int_x^\infty \int_y^\infty (1-F(z))dzdy + \frac{\lambda}{\pi_2} b_\infty(0) \int_x^\infty (1-F(y))dy.$$

Next, using the defining equation (1.4.7) for δ it is readily verified that

$$\int_0^\infty e^{\delta y} w(y)dy = \frac{1}{\delta^2}$$

and hence

$$(1.4.32)\qquad \lim_{x\to\infty} e^{\delta x} b_\infty(x) = \frac{1}{\delta^2 \nu}.$$

This result could alternatively be obtained from the asymptotic behaviour of $q_\infty(x)$ given by (1.4.6). We first note that equations (1.4.13) and (1.4.20) with $b_\infty(x)$ substituted for b(x) have unique solutions. This holds because of the fact that either of these two equations can be rewritten into a defective renewal equation. Differentiating equation (1.4.20) yields

$$(1.4.33)\qquad b_\infty''(x) = \frac{\lambda}{\pi_2-\lambda E[D]} (1-F(x)) + \frac{\lambda}{\pi_2} b_\infty'(x) - \frac{\lambda}{\pi_2} \int_0^x b_\infty'(x-y)dF(y).$$

By comparing equation (1.4.33) with equation (1.4.13) we find that $-\pi_2 q_\infty(x)/(\pi_2-\lambda E[D])$ is a solution of (1.4.33). This solution must be unique. Since (1.4.33) is obtained by differentiation of (1.4.20) we find

$$(1.4.34)\qquad b_\infty(x) = \frac{\pi_2}{\pi_2-\lambda E[D]} \int_x^\infty q_\infty(y)dy$$

is the unique solution of (1.4.20). Then the equations (1.4.26), (1.4.30) and (1.4.32) follow from (1.4.6), (1.4.12) and (1.4.34).

Equation (1.4.34) suggests to use again an approximating function that is a mixture of two exponentials. We choose this mixture such that the

conditions given by (1.4.30), (1.4.32) and $b_\infty'(0)=-\lambda E[D]/(\pi_2-\lambda E[D])$ are met. Hence we propose the following

Approximation 1.4.2. For all $0\leq x\leq M$

$$b(x) = b_\infty(x) - b_\infty(M).$$

with

$$b_\infty(x) \cong \frac{\pi_2\alpha}{\beta(\pi_2-\lambda E[D])}\, e^{-\beta x} + \frac{1}{\delta^2\nu}\, e^{-\delta x}, \quad x \geq 0,$$

where the constants α and β are given by (1.4.17) and (1.4.18) respectively. The equation (1.4.19) and approximation 1.4.2 together yield the desired expressions for $b(x)$.

It remains to find an approximation for $c(x)$, i.e. the expected total cumulative backlog at the epoch at which the inventory level reaches the value M, given $X(0)=x\leq M$. Again we first consider the case $x\leq 0$. By using the lack of memory of the Poisson arrival process we find

$$(1.4.35) \qquad c(x) = E[C(T_0(x))] + c(0), \quad x \leq 0.$$

It is easy to see that $\{-X(t),\ 0\leq t\leq T_0(x)\}$ corresponds to the workload process in the following M/G/1-type queueing system. Jobs arrive according to a Poisson process with rate λ and the amounts of work involved by the jobs are independent random variables with common distribution function F. Also, work is processed at a constant rate of π_2 per unit time whenever there is work in the system. Here we assume that the initial workload equals $-x$ and we only observe the process until the first epoch at which the system becomes empty. If we assume that a holding cost at rate w is incurred if the workload equals w it follows from results in Tijms [1977] that

$$\text{expected holding cost incurred until the system is empty} = \frac{(-x)^2}{2(\pi_2-\lambda E[D])} + \frac{E[D^2](-x)}{2(\pi_2-\lambda E[D])^2}, \quad (-x) \geq 0.$$

Using the cost interpretation of the cumulative backlog given in section 1.2 we thus obtain

$$E[C(T_0(x))] = \frac{(-x)^2}{2(\pi_2-\lambda E[D])} + \frac{\lambda E[D^2](-x)}{2(\pi_2-\lambda E[D])^2}, \quad x \leq 0.$$

Hence it follows from (1.4.35) that

$$c(x) = \frac{x^2}{2(\pi_2-\lambda E[D])} - \frac{\lambda E[D^2]x}{2(\pi_2-\lambda E[D])^2} + c(0), \quad x \leq 0. \tag{1.4.36}$$

Using the lack of memory of the Poisson arrival process as done in the derivation of (1.4.25) and defining $c_\infty(x)$ in the obvious way, we obtain

$$c(x) = c_\infty(x) - c_\infty(M), \quad x \leq M. \tag{1.4.37}$$

By conditioning on the possible events in the interval $(0,\Delta x/\pi_2)$ where Δx is small we derive the following integro-differential equation,

$$c_\infty'(x) = \frac{-\lambda}{2\pi_2(\pi_2-\lambda E[D])} \int_x^\infty (y-x)^2 dF(y) - \frac{\lambda^2 E[D^2]}{2\pi_2(\pi_2-\lambda E[D])^2}$$

$$\int_x^\infty (y-x)dF(y) - \frac{\lambda}{\pi_2} c_\infty(0)(1-F(x)) + \frac{\lambda}{\pi_2} c_\infty(x) - \frac{\lambda}{\pi_2} \int_0^x c_\infty(x-y)dF(y), \tag{1.4.38}$$

$$x \geq 0.$$

Differentiation of (1.4.38) yields an equation that equals $-1/\pi_2$ times equation (1.4.20). Using the fact that (1.4.38) has a unique solution we obtain

$$c_\infty(x) = \frac{\lambda}{\pi_2} \int_x^\infty b_\infty(y)dy. \tag{1.4.39}$$

In principle we are now able to give an approximation for $c(x)$, $x \geq 0$, based on the equations (1.4.37) and (1.4.39) and using approximation (1.4.2). However, by doing so, it appears that in general $c_\infty(0)$ is not exactly matched. By integration of both sides of (1.4.38) we obtain

$$c_\infty(x) = v(x) + \int_0^x c_\infty(x-y) \frac{\lambda}{\pi_2} (1-F(y))dy \tag{1.4.40}$$

with

$$(1.4.41)\qquad v(x) = c_\infty(0) - \frac{\lambda E[D^3]}{6\pi_2(\pi_2-\lambda E[D])} - \frac{\lambda^2(E[D^2])^2}{4\pi_2(\pi_2-\lambda E[D])^2} - \frac{\lambda E[D]}{\pi_2}\, c_\infty(0)$$

$$+ \frac{\lambda}{6\pi_2(\pi_2-\lambda E[D])} \int_x^\infty (y-x)^3 dF(y) + \frac{\lambda^2 E[D^2]}{4\pi_2(\pi_2-\lambda E[D])^2} \cdot$$

$$\cdot \int_x^\infty (y-x)^2 dF(y) + \frac{\lambda}{\pi_2}\, c_\infty(0) \int_x^\infty (1-F(y))dy.$$

The equation (1.4.40) is again a defective renewal equation. Hence $\lim_{x\to\infty} c_\infty(x) = \lim_{x\to\infty} v(x)/(1-\lambda E[D]/\pi_2)$ and using $\lim_{x\to\infty} c_\infty(x)=0$ we find from (1.4.41)

$$(1.4.42)\qquad c_\infty(0) = \frac{\lambda E[D^3]}{6(\pi_2-\lambda E[D])^2} + \frac{\lambda^2(E[D^2])^2}{4(\pi_2-\lambda E[D])^3} \cdot$$

It is easy to show that for all $a\in\mathbb{R}^+$ and for any f

$$\lim_{x\to\infty} e^{ax} f(x) = b \Rightarrow \lim_{x\to\infty} e^{ax} \int_x^\infty f(y)dy = \frac{b}{a}.$$

Hence it follows from (1.4.32) and (1.4.39) that

$$(1.4.43)\qquad \lim_{x\to\infty} e^{\delta x} c_\infty(x) = \frac{1}{\pi_2 \delta^3 \nu} \cdot$$

This result could alternatively be derived by multiplying equation (1.4.40) by $e^{\delta x}$ and applying the Key Renewal Theorem.

Invoking (1.4.42) and (1.4.43) the approximation resulting from equation (1.4.37) and approximation 1.4.2 yields an exact expression for $c_\infty(0)$ only if

$$\frac{\lambda E[D^3]}{6(\pi_2-\lambda E[D])^2} + \frac{\lambda^2(E[D^2])^2}{4(\pi_2-\lambda E[D])^3} = \frac{\alpha}{\beta^2(\pi_2-\lambda E[D])} + \frac{1}{\pi_2 \delta^3 \nu} \cdot$$

This equation is not true in general. Therefore we determine new constants α' and β' acting as α and β, respectively, such that the resulting approximation holds exactly for x=0. Thus we propose

Approximation 1.4.3. For all $0\le x\le M$

$$c(x) = c_\infty(x) - c_\infty(M)$$

with

$$c_\infty(x) \cong \alpha' e^{-\beta' x} + \frac{1}{\pi_2 \delta^3 \nu} e^{-\delta x}, \quad x \geq 0$$

and

$$\alpha' := \frac{\lambda E[D^3]}{6(\pi_2 - \lambda E[D])^2} + \frac{\lambda^2 (E[D^2])^2}{4(\pi_2 - \lambda E[D])^3} - \frac{1}{\pi_2 \delta^3 \nu} \tag{1.4.44}$$

$$\beta' := \left[\frac{\lambda E[D^2]}{2(\pi_2 - \lambda E[D])^2} - \frac{1}{\pi_2 \delta^2 \nu} \right] / \alpha'. \tag{1.4.45}$$

The equation (1.4.36) and approximation 1.4.3 together determine the desired expression for $c(x)$.

As for approximation 1.4.1 we need that $\beta' > \delta$, otherwise equation (1.4.43) would be violated by the approximate $c_\infty(x)$. It turns out that there are cases where $\beta' \leq \delta$, whereas $\beta > \delta$. Therefore we suggest the following rule of thumb,

If $\beta' \leq \delta$ then replace β' by β in approximation 1.4.3.

This recipe proved to give satisfactory results with regard to the accuracy of the approximation for $c(x)$.

We conclude this section by a number of remarks.

Remark 1.4.1. All approximations in this section are exact in case F is a K_2-distribution. This follows by combining the relations (1.4.2), (1.4.25), (1.4.34), (1.4.37) and (1.4.39) with the fact that the approximation for $q_\infty(x)$ is exact.

Remark 1.4.2. We have given approximations that are sums of two exponential functions. The coefficients involved were carefully chosen to match both the asymptotic behaviour at infinity and the behaviour near the origin in a simple way. We mentioned that there are several other possible ways to achieve the same. In the paper by De Kok et al [1984] the following coupling idea is used which yields very good approximations as well. Suppose we want to approximate an unknown function f. We know the following properties of f:

$$f(0) = a_0, \quad \lim_{x\to\infty} e^{\delta x} f(x) = a_\infty.$$

Then the approximation $\tilde{f}$ of f suggested in De Kok et al [1984] is determined by choosing some positive number x_0 and defining $\tilde{f}$ by

$$\tilde{f}(x) = \begin{cases} a_0 e^{-\beta x}, & 0 \leq x \leq x_0 \\ a_\infty e^{-\delta x}, & x \geq x_0 \end{cases}$$

where

$$\beta = \delta + \frac{1}{x_0} \ln [a_0/a_\infty].$$

The choice of β is based on the continuity of $\tilde{f}$ at x_0. Because we are free to choose x_0 we can try to find that value of x_0 yielding a robust approximation. For the functions $q_\infty(x)$, $b_\infty(x)$ and $c_\infty(x)$ De Kok et al [1984] propose

$$x_0 = E[D].$$

The advantage of this method is that it is always possible to choose x_0 such that $q_\infty(x)$, $b_\infty(x)$ and $c_\infty(x)$ are strictly decreasing for $x \geq 0$. This approximation is exact only for the case of exponential demand.

Remark 1.4.3. A consequence of the assumptions for the model discussed in this chapter is that the arrival process is not influenced by the inventory level. In practical situations it may happen that some customers are discouraged because of a negative inventory level or equivalently because of having to wait for the fulfillment of demand. This phenomenon can be modelled in the following way. Customers arrive according to a Poisson process with rate

$$\begin{cases} \lambda & \text{while } X(t) \geq 0 \\ \tilde{\lambda} & \text{while } X(t) < 0 \end{cases},$$

with $\tilde{\lambda}<\lambda$. The demands of customers are independent random variables with common distribution function F. Excess demand is backlogged. We again assume that the inventory is controlled by an (m,M)-rule.

It is obvious that the analysis for this model only differs with respect to the basic functions associated with production rate π_2. Partly we can proceed along the same lines as in this section, partly the analysis is based on ideas that will be clarified in subsequent chapters.
Yet we think it appropriate to summarize the results here. We use a "~" to mark the basic functions and the random variables associated with the new model. Then we have the following results.

$$\tilde{q}(x) = q(x) \qquad x \geq 0. \tag{1.4.46}$$

$$\tilde{b}(x) = \begin{cases} (\pi_2-\lambda E[D])(\pi_2-\tilde{\lambda}E[D])^{-1}b(x) & x \geq 0 \\ -\tilde{\lambda}E[D](\pi_2-\tilde{\lambda}E[D])^{-1}x + (\pi_2-\lambda E[D])(\pi_2-\tilde{\lambda}E[D])^{-1}b(0), & x \leq 0 \end{cases} \tag{1.4.47}$$

$$\tilde{t}_2(x) = \begin{cases} (M-x)(\pi_2-\lambda E[D])^{-1}-(\lambda-\tilde{\lambda})E[D][\pi_2(\pi_2-\tilde{\lambda}E[D])]^{-1}b(x) & x \geq 0 \\ (-x)/\pi_2-\tilde{\lambda}E[D])^{-1}+M(\pi_2-\lambda E[D])^{-1} & \\ \qquad -(\lambda-\tilde{\lambda})E[D][\pi_2(\pi_2-\tilde{\lambda}E[D])]^{-1}b(0) & x \leq 0 \end{cases} \tag{1.4.48}$$

$$\tilde{c}(x) = \tilde{c}_\infty(x) - \tilde{c}_\infty(M), \qquad x \leq M. \tag{1.4.49}$$

$$\tilde{c}_\infty(0) = \frac{\lambda E[D^3]}{6(\pi_2-\tilde{\lambda}E[D])(\pi_2-\lambda E[D])} + \frac{\lambda\tilde{\lambda}'(E[D^2])^2}{4(\pi_2-\tilde{\lambda}E[D])^2(\pi_2-\lambda E[D])}. \tag{1.4.50}$$

$$\tilde{c}_\infty(x) = \frac{x^2}{2(\pi_2-\tilde{\lambda}E[D])} + \frac{\tilde{\lambda}E[D^2](-x)}{2(\pi_2-\tilde{\lambda}E[D])^2} + \tilde{c}_\infty(0), \qquad x \leq 0. \tag{1.4.51}$$

$$\lim_{x\to\infty} e^{\delta x}\tilde{c}_\infty(x) = \frac{1}{\nu\delta^2(\pi_2-\tilde{\lambda}E[D])}\left\{\frac{\pi_2-\lambda E[D]}{\pi_2\delta} - \frac{(\lambda-\tilde{\lambda}')E[D^2]}{2(\pi_2-\tilde{\lambda}E[D])}\right\}. \tag{1.4.52}$$

The equations (1.4.50) and (1.4.52) can be used to find an approximation for $\tilde{c}_\infty(x)$, $x\geq 0$, that is a mixture of two exponentials. Again the equations (1.2.8)-(1.2.10) carry through for this model.

So we can compute expressions for $E[\tilde{T}]$, $E[\tilde{B}]$, $E[\tilde{Q}]$ and $E[\tilde{C}]$ from the results derived in section 1.3 and (1.4.46)-(1.4.52). An expression for $E[\tilde{S}]$ can be derived from

$$\frac{E[\tilde{S}]}{E[\tilde{N}]} = \frac{E[\tilde{Q}]}{E[\tilde{N}]} + \frac{E[\tilde{J}]}{E[\tilde{T}]} . \tag{1.4.53}$$

$$E[\tilde{J}] = E[\tilde{B}]/\pi_2, \tag{1.4.54}$$

$$E[\tilde{N}] = \lambda(E[\tilde{T}]-E[\tilde{J}]) + \tilde{\lambda}E[\tilde{J}]. \tag{1.4.55}$$

A special case of this model is that in which customers arrive according to a Poisson process and a customer, who has to wait for the fulfillment of his demand leaves directly with probability $0<p\leq 1$ without taking any inventory. This situation can be modelled by taking $\tilde{\lambda}=\lambda(1-p)$. In particular the case of $\tilde{\lambda}=0$, corresponding to the model in which customers renege if they have to wait for the fulfillment of their demand, is a special case of the model that is discussed in chapter 4.

1.5. Numerical results and conclusions.

In this section we give numerical results validating the accuracy of the approximations and we discuss the sensitivity of the switch-over level m to more than the first two moments of the demand distribution. Our extensive numerical investigations indicate that in general the approximations are quite accurate and thus are suited for use in practice. Also, it turns out that in general two-moment approximations for the switching level m are justified provided the coefficient of variation of the individual demand is not too large. Some rules of thumb for the use of the approximations are given at the end of this section.

As stated in section 1.1 we sequentially determine M-m and m. We first determine the value of M-m by using cost considerations only and next we determine the value of m by invoking the service level constraint. The value of M-m is determined as follows. We assume that a switch-over cost K is incurred each time the production rate is switched from π_1 to π_2. Also, a holding cost at rate h.x is incurred when the on-hand inventory equals $x\geq 0$. The constants K and h are positive. Under this cost-structure we suggest the following formula for M-m,

$$(1.5.1) \qquad M-m = \left\{\frac{2K(\pi_2-\lambda E[D])(\lambda E[D]-\pi_1)}{h(\pi_2-\pi_1)}\right\}^{\frac{1}{2}}.$$

This formula generalizes the classical economic order quantity formula from inventory theory with instantaneous delivery of ordered items, cf. Hadley and Whitin [1963]. To motivate the formula (1.5.1), we note that it is empirically known that the uncertainty of demand mainly influences the reorder point rather than the order quantity. Thus a reasonable approximation for the order quantity may be obtained by replacing the stochastic demand process by its average value and thereby considering a deterministic inventory system. Therefore consider now the deterministic production-inventory problem in which the demand occurs at a constant rate $\lambda E[D]$. In this deterministic system it follows that under the (m,M)-rule the length of a cycle equals $(M-m)/(\lambda E[D]-\pi_1)+(M-m)/(\pi_2-\lambda E[D])$ and the average on-hand inventory equals $m+\frac{1}{2}(M-m)$ so that the average holding and switch-over costs per unit time are given by

$$(1.5.2) \qquad \frac{K(\lambda E[D]-\pi_1)(\pi_2-\lambda E[D])}{(\pi_2-\pi_1)(M-m)} + \frac{h}{2}(M-m) + hm.$$

By minimizing this expression with respect to M-m, we get the formula (1.5.1).

In chapter 5 it is shown that the performance of the production-quantity formula (1.5.1) is indeed very good with respect to the minimization of the long-run average costs per unit time. However, the expression (1.5.2) may predict rather unsatisfactorily the actual average costs.

In testing the accuracy of the approximations we consider the β-service measure, associated with the fraction of demand satisfied directly from stock on hand, and the γ-service measure associated with the fraction of customers, whose demands are met directly from stock on hand. In the sensitivity analysis we present results only for the β-service measure, but the conclusions for the other service measures are very much the same.

In all examples considered in tables 1.5.1-1.5.3 M-m is predetermined by the formula (1.5.1), in which h=1 and K=25. We do not consider the case of M=m (i.e. K=0), since in that case we have an exact expression for p(x,u). Hence accuracy for the case of M=m is implied by accuracy for the case of M>m. Conclusions drawn concerning sensitivity for the case of M>m also apply to the case of M=m. Next, using our approximate results obtained

in sections 1.3 and 1.4, we determine the switch-over level m such that the required service level is achieved. The parameters are varied as follows. The production rate π_1 has the three values -0.5, 0 and 0.5, the production rate π_2 has the three values 1.25, 2 and 5. The service levels β and γ are varied as 0.95 and 0.99. In all examples we take $\lambda=1$ and $E[D]=1$. Let $c_D=\sigma(D)/E[D]$ denote the coefficient of variation of the demand size D.

In tables 1.5.1 and 1.5.2 we vary c_D^2 as 0, 1/3, 2/3 and 2. We consider the following demand distributions,

(i) deterministic demand ($c_D^2=0$)
(ii) gamma demand ($c_D^2=1/3$, 2/3, 2).

In practical inventory applications the gamma distribution gives very often an excellent fit to the empirical demand distribution. The gamma distribution is completely specified by its first two moments and it can achieve any positive value of the coefficient of variation c_D of the demand size D.

In tables 1.5.1 and 1.5.2 we give the approximate (m,M)-rules and their actual values β_{act} and γ_{act}, respectively, of the achieved service levels. For the cases with both $c_D^2=0$ and $\pi_1=0$ we cannot apply our approximations for p(M-m,u) and $t_1(M-m)$, because the demand D is arithmetic. For these cases we have used the exact formulas for p(M-m,u) and $t_1(M-m)$, which are trivial to derive. The actual service levels β_{act} and γ_{act} are determined by computer simulation. In each example we have simulated 250,000 customers. The notation 0.953(2) for β_{act} means that the 95% confidence interval of the simulated value is given by 0.951-0.955. If the demand is deterministic then the fraction of customers, whose demands cannot be met directly from stock on hand, is a discontinuous function of the switching level m with fixed M-m. Hence some service levels cannot be achieved. We found that for the case of $\pi_1=0$ and $\pi_2=5$ the γ-service level 0.95 could not be achieved.

In table 1.5.3 we deal with the sensitivity of the switch-over level m to more than the first two moments of the demand distribution. As stated above the gamma distribution is of great importance for practical applications. Unfortunately the computations of the approximate (m,M)-rules require some special numerical procedures for computing incomplete gamma integrals and numerical integration. However, it is always possible to find a distribution which has the same first or three moments as the gamma

Table 1.5.1. The approximate (m,M)-rules and their actual β-service levels.

			$c_D^2=0$			$c_D^2=1/3$		
π_1	π_2	β	m	M	β_{act}	m	M	β_{act}
-0.5	1.25	0.95	5.57	8.84	0.950(6)	8.09	11.37	0.953(4)
-0.5	2	0.95	1.02	6.49	0.950(3)	1.82	7.30	0.950(3)
-0.5	5	0.95	0.26	7.65	0.950(1)	0.57	7.96	0.950(2)
0	1.25	0.95	5.74	8.90	0.952(5)	8.05	11.21	0.947(6)
0	2	0.95	1.44	6.44	0.949(2)	1.87	6.87	0.951(3)
0	5	0.95	0.45	6.78	0.950(1)	0.65	6.98	0.950(2)
0.5	1.25	0.95	5.15	8.03	0.952(5)	7.52	10.41	0.952(6)
0.5	2	0.95	0.76	4.84	0.948(2)	1.51	5.59	0.949(3)
0.5	5	0.95	0.04	4.75	0.950(1)	0.33	5.05	0.950(1)
-0.5	1.25	0.99	9.31	12.58	0.991(3)	13.26	16.54	0.991(3)
-0.5	2	0.99	2.30	7.78	0.991(2)	3.74	9.22	0.990(2)
-0.5	5	0.99	0.86	8.24	0.990(1)	1.65	9.04	0.990(1)
0	1.25	0.99	9.47	12.64	0.991(2)	13.22	16.38	0.990(3)
0	2	0.99	2.73	7.73	0.990(1)	3.79	8.79	0.989(2)
0	5	0.99	0.96	7.28	0.989(1)	1.72	8.05	0.990(1)
0.5	1.25	0.99	8.88	11.77	0.992(3)	12.69	15.58	0.990(4)
0.5	2	0.99	2.05	6.13	0.990(2)	3.43	7.51	0.990(2)
0.5	5	0.99	0.66	5.37	0.989(1)	1.41	6.12	0.990(1)

			$c_D^2=2/3$			$c_D^2=2$		
π_1	π_2	β	m	M	β_{act}	m	M	β_{act}
-0.5	1.25	0.95	10.65	13.92	0.953(7)	21.02	24.29	0.951(8)
-0.5	2	0.95	2.72	8.19	0.950(3)	6.71	12.19	0.949(4)
-0.5	5	0.95	0.96	8.35	0.949(2)	3.01	10.40	0.948(3)
0	1.25	0.95	10.58	13.74	0.947(8)	20.85	24.01	0.958(9)
0	2	0.95	2.77	7.77	0.949(3)	6.71	11.71	0.949(4)
0	5	0.95	1.06	7.39	0.949(2)	3.16	9.49	0.949(2)
0.5	1.25	0.95	9.92	12.81	0.950(7)	19.68	22.56	0.954(9)
0.5	2	0.95	2.33	6.41	0.950(3)	5.88	9.96	0.950(5)
0.5	5	0.95	0.69	5.41	0.949(2)	2.53	7.25	0.949(2)
-0.5	1.25	0.99	17.26	20.53	0.988(6)	33.39	36.67	0.992(4)
-0.5	2	0.99	5.28	10.76	0.990(1)	11.91	17.39	0.990(3)
-0.5	5	0.99	2.51	9.89	0.990(1)	6.43	13.82	0.990(2)
0	1.25	0.99	17.19	20.35	0.990(4)	33.23	36.39	0.988(6)
0	2	0.99	5.33	10.33	0.989(2)	11.92	16.92	0.989(3)
0	5	0.99	2.61	8.93	0.990(1)	6.58	12.91	0.990(2)
0.5	1.25	0.99	16.53	19.42	0.993(3)	32.05	34.94	0.992(4)
0.5	2	0.99	4.89	8.97	0.989(2)	11.08	15.17	0.989(2)
0.5	5	0.99	2.23	6.95	0.990(1)	5.95	10.67	0.989(2)

Table 1.5.2. The approximate (m,M)-rules and their actual γ-service levels.

			$c_D^2=0$			$c_D^2=1/3$		
π_1	π_2	γ	m	M	γ_{act}	m	M	γ_{act}
-0.5	1.25	0.95	6.05	9.33	0.950(6)	8.43	11.71	0.955(6)
-0.5	2	0.95	1.47	6.95	0.950(3)	2.17	7.65	0.949(2)
-0.5	5	0.95	0.73	8.12	0.950(2)	0.93	8.32	0.950(2)
0	1.25	0.95	6.22	9.38	0.952(6)	8.39	11.55	0.951(5)
0	2	0.95	1.89	6.89	0.949(2)	2.22	7.22	0.948(3)
0	5	0.95	0.68	7.00	0.963(2)	1.01	7.34	0.950(2)
0.5	1.25	0.95	5.63	8.51	0.952(6)	7.86	10.75	0.952(7)
0.5	2	0.95	1.21	5.29	0.948(2)	1.86	5.94	0.952(3)
0.5	5	0.95	0.50	5.21	0.951(2)	0.69	5.40	0.951(2)
-0.5	1.25	0.99	9.79	13.06	0.991(3)	13.60	16.88	0.991(3)
-0.5	2	0.99	2.75	8.23	0.991(2)	4.09	9.56	0.990(2)
-0.5	5	0.99	1.42	8.63	0.990(1)	2.01	9.40	0.989(1)
0	1.25	0.99	9.96	13.12	0.991(3)	13.56	16.72	0.991(3)
0	2	0.99	3.18	8.18	0.990(1)	4.14	9.14	0.989(2)
0	5	0.99	1.37	7.69	0.987(1)	2.08	8.41	0.990(1)
0.5	1.25	0.99	9.36	12.25	0.992(2)	13.03	15.92	0.993(3)
0.5	2	0.99	2.50	6.58	0.990(2)	3.78	7.86	0.990(2)
0.5	5	0.99	1.01	5.73	0.989(1)	1.77	6.48	0.990(1)

			$c_D^2=2/3$			$c_D^2=2$		
π_1	π_2	γ	m	M	γ_{act}	m	M	γ_{act}
-0.5	1.25	0.95	10.82	14.07	0.952(6)	20.46	23.74	0.955(7)
-0.5	2	0.95	2.91	8.39	0.949(4)	8.03	11.51	0.950(4)
-0.5	5	0.95	1.17	8.55	0.950(2)	2.22	9.61	0.949(2)
0	1.25	0.95	10.75	13.92	0.948(8)	20.30	23.46	0.945(7)
0	2	0.95	2.96	7.96	0.947(3)	6.03	11.03	0.951(4)
0	5	0.95	1.27	7.59	0.950(2)	2.36	8.69	0.950(2)
0.5	1.25	0.95	10.10	12.99	0.950(8)	19.12	22.01	0.944(7)
0.5	2	0.95	2.52	6.60	0.950(3)	5.20	9.28	0.950(4)
0.5	5	0.95	0.90	5.61	0.950(2)	1.73	6.44	0.950(2)
-0.5	1.25	0.99	17.43	20.70	0.991(4)	32.84	36.11	0.993(3)
-0.5	2	0.99	5.47	10.95	0.990(2)	11.23	16.71	0.990(2)
-0.5	5	0.99	2.72	10.11	0.990(1)	5.56	12.95	0.989(1)
0	1.25	0.99	17.36	20.52	0.989(4)	32.67	35.83	0.990(4)
0	2	0.99	5.52	10.52	0.990(2)	11.23	16.23	0.990(2)
0	5	0.99	2.82	9.14	0.990(1)	5.72	12.04	0.990(2)
0.5	1.25	0.99	16.71	19.59	0.992(4)	31.49	34.38	0.990(4)
0.5	2	0.99	5.08	9.16	0.989(2)	10.40	14.48	0.990(3)
0.5	5	0.99	2.45	7.16	0.990(1)	5.09	9.80	0.990(1)

distribution and which is easy to use from a computational point of view.

If $0<c_D^2<1$ and so $1/k\leqq c_D^2<1/(k-1)$ for some integer $k\geqq 2$, then we can match the first two moments of the demand D by a unique mixture

$$f(x) = p\mu^{k-1}\frac{x^{k-2}}{(k-2)!}e^{-\mu x} + (1-p)\mu^{k}\frac{x^{k-1}}{(k-1)!}e^{-\mu x}, \quad x\geqq 0$$

of E_{k-1} and E_k demand densities with the same scale parameters, where

$$p = \frac{1}{1+c_D^2}\left[kc_D^2-\{k(1+c_D^2)-k^2c_D^2\}^{\frac{1}{2}}\right], \quad \mu = \frac{k-p}{E[D]}.$$

This mixture of E_{k-1} and E_k densities has always a unimodal shape like the gamma density. We note that this is <u>not</u> true for a mixture of E_1 and E_k densities with the same scale parameters which mixtures can be used also to match the first two moments of D.

For the case of $c_D^2>\frac{1}{2}$, a useful class of demand distributions is the class of K_2-distributions, which has already been described in section 1.4. We recall that the Laplace-Stieltjes transform of a K_2-distribution has the form $(1+a_0s)/(1+a_1s+a_2s^2)$. If the demand size D is gamma distributed with $c_D^2>\frac{1}{2}$, then there exists a unique K_2-distribution having the same first three moments as D. This can be verified by using that a gamma distributed random variable D has the property $E[D^3]E[D]\geqq(\leqq)1\frac{1}{2}(E[D^2])^2$ if $c_D^2\geqq(\leqq)1$; see Whitt [1982] for further details. The parameters a_0, a_1 and a_2 of the K_2-distribution having the same first three moments as the gamma distributed demand size D with $c_D^2>\frac{1}{2}$ are given by

$$a_0 = a_1-E[D], \quad a_1 = \frac{2E[D^2]}{3E[D]}, \quad a_2 = a_1E[D]-\tfrac{1}{2}E[D^2].$$

To give the probability density f(x) of this K_2-distribution (cf. also Cox [1955]), let

$$b_1 = \frac{2}{E[D]} + \frac{2}{E[D]}\left\{\frac{c_D^2-\frac{1}{2}}{c_D^2+1}\right\}^{\frac{1}{2}}, \quad b_2 = \frac{4}{E[D]} - b_1,$$

and

$$q = b_1(b_2E[D]-1)/(b_2-b_1).$$

Table 1.5.3. Sensitivity analysis for the switch-over level m.

			c_D^2=0.4			c_D^2=0.8		
π_1	π_2	β	gamma	$E_{2,3}$	$E_{1,3}$	gamma	$E_{1,2}$	$E_{1,3}$
-0.5	1.25	0.95	8.603	8.594	8.580	11.676	11.634	11.564
-0.5	2	0.95	1.996	1.989	1.978	3.092	3.058	2.999
-0.5	5	0.95	0.645	0.643	0.639	1.137	1.124	1.102
0	1.25	0.95	8.552	8.543	8.528	11.596	11.553	11.483
0	2	0.95	2.047	2.040	2.029	3.139	3.104	3.044
0	5	0.95	0.729	0.727	0.722	1.245	1.230	1.204
0.5	1.25	0.95	8.001	7.991	7.974	10.891	10.843	10.763
0.5	2	0.95	1.671	1.663	1.649	2.665	2.625	2.556
0.5	5	0.95	0.398	0.395	0.387	0.852	0.829	0.792
-0.5	1.25	0.99	14.059	14.043	14.017	18.859	18.782	18.655
-0.5	2	0.99	4.040	4.024	4.000	5.917	5.836	5.703
-0.5	5	0.99	1.815	1.801	1.780	2.869	2.799	2.678
0	1.25	0.99	14.008	13.992	13.966	18.779	18.701	18.573
0	2	0.99	4.091	4.075	4.050	5.964	5.882	5.749
0	5	0.99	1.895	1.879	1.857	2.975	2.902	2.774
0.5	1.25	0.99	13.457	13.440	13.411	18.074	17.991	17.854
0.5	2	0.99	3.715	3.698	3.671	5.489	5.403	5.261
0.5	5	0.99	1.566	1.549	1.525	2.581	2.501	2.363
			c_D^2=1.5			c_D^2=3		
π_1	π_2	β	gamma	2-mom.	3-mom.	gamma	2-mom.	3-mom.
-0.5	1.25	0.95	17.110	17.502	17.105	28.894	30.460	28.859
-0.5	2	0.95	5.162	5.502	5.162	9.906	11.376	9.896
-0.5	5	0.95	2.178	2.370	2.195	4.836	6.121	4.915
0	1.25	0.95	16.979	17.382	16.974	28.649	30.264	28.612
0	2	0.95	5.187	5.534	5.187	9.852	11.353	9.840
0	5	0.95	2.314	2.526	2.329	4.989	6.305	5.044
0.5	1.25	0.95	16.001	16.452	15.996	27.080	28.887	27.040
0.5	2	0.95	4.511	4.900	4.512	8.695	10.344	8.683
0.5	5	0.95	1.790	2.058	1.803	4.130	5.582	4.161
-0.5	1.25	0.99	27.319	28.036	27.305	45.597	48.323	45.508
-0.5	2	0.99	9.372	10.139	9.340	17.111	19.803	16.915
-0.5	5	0.99	4.892	5.609	4.877	9.654	12.319	9.545
0	1.25	0.99	27.189	27.915	27.174	45.352	48.127	45.261
0	2	0.99	9.397	10.172	9.365	17.056	19.780	16.858
0	5	0.99	5.031	5.773	5.011	9.812	12.503	9.675
0.5	1.25	0.99	26.210	26.986	26.197	43.783	46.750	43.689
0.5	2	0.99	8.720	9.537	8.690	15.897	18.771	15.702
0.5	5	0.99	4.508	5.308	4.485	8.958	11.780	8.791

Note that $0 \leq q \leq 1$ if $c_D^2 \geq 1$ and $q<0$ otherwise. The associated density f(x) has the form

$$f(x) = p_1\mu_1 e^{-\mu_1 x} + p_2\mu_2 e^{-\mu_2 x}$$

with $p_1=q$, $p_2=1-q$, $\mu_1=b_1$ and $\mu_2=b_2$. Since $0 \leq q \leq 1$ for $c_D^2 \geq 1$, we have that for $c_D^2 \geq 1$ the K_2-density belongs to the well-known class of hyperexponential densities of order two (H_2-density). The H_2-density is always unimodal with a maximum at $x=0$. The H_2-density is not uniquely determined by its first two moments. We have seen above that a unique H_2-density can be found having the same first three moments as a gamma density with $c_D^2 \geq 1$. Another H_2-density that is often used in practice is the H_2 density with balanced means, i.e. $p_1/\mu_1 = p_2/\mu_2$ (cf. table 1.3.1). The parameters of this H_2 density are given by

$$p_1 = \tfrac{1}{2}\left[1 + \left\{\frac{c_D^2-1}{c_D^2+1}\right\}^{\frac{1}{2}}\right], \quad p_2 = 1-p_1, \quad \mu_1 = \frac{2p_1}{E[D]}, \quad \mu_2 = \frac{2p_2}{E[D]}.$$

In table 1.5.3 we give the switch-over level m for several demand distributions having the same first two or three moments. We assume gamma distributed demand D with $c_D^2=0.4$, 0.8, 1.5 and 3. For $c_D^2=0.4$ and 0.8 we also consider both unimodal mixtures of E_{k-1} and E_k demand densities ($E_{k-1,k}$) with the same scale parameters and mixtures of E_1 and E_k demand densities ($E_{1,k}$) with the same scale parameters. For $c_D^2=1.5$ and 3 we also consider both the H_2-density with balanced means and having the same first two moments as the gamma demand D and the H_2-density having the same first three moments as the gamma demand D

Conclusions.

From our numerical investigations we draw some conclusions concerning the accuracy of the approximations and the sensitivity of the level m.

(i) Accuracy. In general the approximations yield excellent results provided M-m=0 or M-m satisfies condition 1.3.1. Some care should be taken in applying the approximations when both $\lambda E[D]/\pi_2$ is small and the required service level is low. It is a well-known phenomenon that the performance of many approximations deteriorates for very light traffic, i.e. when $\lambda E[D]/\pi_2$ gets small. To give some indication as to when the approximations can be applied, we claim that a sufficient accuracy is guaranteed when

$\lambda E[D]/\pi_2 \geq 0.1$ and the required service level ≥ 0.9.

This condition holds obviously in practical applications.

(ii) Sensitivity. As might be expected, the switch-over level m becomes increasingly sensitive to more than the first two moments of the demand when c_D^2 gets larger. However, for the practically important case of $0<c_D^2\leq 1$ we found that the m-value depends on the demand distribution F mainly through the first two moments. Here we add the condition that the demand density should satisfy a "reasonable" shape constraint. It is always possible to construct extremal two-point distributions (cf. Whitt [1984]) such that the switch-over level m is quite sensitive to more than the first two moments of these distributions.

As $\lambda E[D]/\pi_2$ gets close to 1, then the m-value gets less sensitive to more than the first two moments. This finding is also familiar from queueing theory (heavy-traffic results).

We recall that a necessary condition for applying the approximations 1.4.1-1.4.3 is that F has an exponentially decreasing tail. This condition is not satisfied if F is a lognormal distribution. Nevertheless the approximate (m,M)-rules obtained by fitting as above a mixture of E_{k-1} and E_k distributions to the demand D yield acceptable results when the actual demand D is lognormal provided $0<c_D^2\leq 1$, $\lambda E[D]/\pi_2 \geq 0.5$ and the required service level is not higher than 99%. A similar conclusion was found in a related study of De Kok and Tijms [1985a].

Another interesting result we found from our numerical investigations is the approximate relation

$$m \cong (1-c_D^2)m(\text{det}) + c_D^2 m(\text{exp}),$$

provided $0<c_D^2\leq 1$, where m(det) and m(exp) denote the m-value for the respective cases of deterministic and exponential demands. Here we remind that for $M-m \geq E[D]$ and $\pi_1=0$ the switching level m(det) should be computed from the approximations 1.3.1, 1.3.2 and 1.3.3 rather than from the exact expressions for $t_1(M-m)$, $p(M-m,u)$ and $E[U]$. The above relationship has been exploited in related models in which approximations like those derived in section 1.4 are not available; see De Kok and Tijms [1985a,b].

2. THE LOST-SALES MODEL.

In this chapter we assume that excess demand is lost rather than backlogged as in chapter 1. Then any customer, whose total demand cannot be met directly from stock on hand, is satisfied with the amount of inventory available, while the excess demand is lost.

We focus again on finding tractable expressions for a number of service measures of interest, such as the long-run fraction of demand that is lost and the average number of stockout occurrences per unit time. Towards this end we derive relations between the lost-sales model and the backlog model studied in chapter 1.

2.1. Description of the model and service measures.

In this section we describe the model and define the service measures that will be considered. Though the assumptions for this model are almost identical to those for the backlog model, we think it is appropriate to refresh the reader's memory.

Customers arrive according to a Poisson process with rate λ. The demands of the customers are independent random variables having a common distribution function F with F(0)=0. The demands are independent of the arrival process. Excess demand is lost.

The production is governed by one out of two production rates π_1 and π_2 such that

$$\pi_1 < \lambda E[D] < \pi_2, \tag{2.1.1}$$

where the generic random variable D denotes the demand of a customer. The inventory is controlled by an (m,M)-rule. We assume an infinite storage capacity.

An exact renewal-theoretic analysis of this model is contained in Doshi et al [1978], but the results obtained there are in general computationally intractable.

The lost-sales model also arises if we want to describe inventories of a perishable item. The item perishes after a fixed time τ. On the other hand the item is continually added to the inventory at rate π_2. Demands for the item occur according to a compound Poisson process. Examples of perishable items are chemical supplies in a continuously processing plant

and blood stored in a bloodbank. The perishable inventory system is considered by Graves [1982] for both exponential and deterministic demand. It is easy to see that Graves deals with the lost-sales model with $\pi_1=0$ and a $(\pi_2\tau, \pi_2\tau)$-rule for controlling the inventory. This rule reflects the finite life-time of the items.

In the lost-sales model the condition $\pi_2>\lambda E[D]$ is not necessary to assure the existence of an equilibrium distribution for the inventory level. Nevertheless, in practical situations, where a high level of customer service is required, the production rate π_2 must be greater than the demand rate.

Now we fix an (m,M)-rule and define for $t\geq 0$,

$N(t)$:= the number of customers that arrive in $(0,t]$.

$V(t)$:= the total demand in $(0,t]$.

$X(t)$:= the inventory level at time t.

$B(t)$:= the amount of demand in $(0,t]$ that is lost.

$Q(t)$:= the number of stockouts that occur in $(0,t]$.

$S(t)$:= the number of customers arriving in $(0,t]$ whose demands are partially lost.

It is immediately clear that

$$Q(t) = S(t), \quad t \geq 0. \tag{2.1.2}$$

Next we define the service measures.

(i) <u>α-service measure</u>.
the long-run average number of stockouts per unit time,

$$\lim_{t\to\infty} \frac{Q(t)}{t}.$$

(ii) β-service measure.
the long-run fraction of demand that is not met directly from stock on hand (and is lost)

$$\lim_{t\to\infty} \frac{B(t)}{V(t)} .$$

(iii) γ-service measure.
the long-run fraction of arriving customers, whose demands are not met directly from stock on hand (and are partially lost),

$$\lim_{t\to\infty} \frac{S(t)}{N(t)}.$$

Through the assumptions the inventory process is again regenerative. We define:

a cycle := the time elapsed between two consecutive epochs at which the production rate is switched from π_2 to π_1.

Assuming that at epoch 0 a cycle starts, define for the given (m,M)-rule

T := the next epoch at which the production rate is switched from π_2 to π_1,

N := N(T), V := V(T), B := B(T), Q := Q(T), S := S(T).

Then it follows that with probability 1

$$(2.1.3) \qquad \lim_{t\to\infty} \frac{Q(t)}{t} = \frac{E[Q]}{E[T]}, \quad \lim_{t\to\infty} \frac{B(t)}{V(t)} = \frac{E[B]}{E[V]}, \quad \lim_{t\to\infty} \frac{S(t)}{N(t)} = \frac{E[S]}{E[N]}.$$

It follows from (2.1.2) and the relations E[N]= E[T] and E[V]=λE[D]E[T] that it suffices to find expressions for E[T], E[B] and E[Q].

As in chapter 1 we express E[T], E[B] and E[Q] in terms of a number of basic functions. Under the condition that at epoch 0 the inventory level equals x+m, x≧0, and production rate π_1 is used, we define

$t_1(x)$:= the expected time until the inventory level decreases below m.

$$p(x,u) := \begin{cases} \text{the probability of an undershoot greater than u of the level m when the inventory decreases below m for the first time,} & 0 \leq u \leq m \\ \text{the probability of a lost demand greater than u-m when the inventory decreases below m for the first time,} & u \geq m. \end{cases}$$

Note that the definition of $p(x,u)$ for the lost-sales model slightly differs from the one given in section 1.3 for the backlog model. As before $t_1(x)$ and $p(x,u)$ are independent of the switching levels m and M.

Next, assuming that at epoch 0 the inventory equals $0 \leq x \leq M$ and the production rate π_2 is used, we define

$t_2(x)$:= the expected time until the inventory reaches the level M.

$b(x)$:= the expected amount of demand that is lost until the inventory reaches the level M.

$q(x)$:= the probability that a stockout occurs before the inventory reaches the level M.

For ease of notation we have suppressed the dependency on M of the functions $t_2(x)$, $b(x)$ and $q(x)$.

As in section 1.2 we can find expressions for E[T], E[B] and E[Q] in terms of $t_1(x)$, $p(x,u)$, $t_2(x)$, $q(x)$ and $b(x)$,

$$E[T] = t_1(M-m) + \int_0^m t_2(m-u)\,d_u(1-p(M-m,u)) + t_2(0)(1-p(M-m,m)) \tag{2.1.4}$$

$$E[Q] = (1-q(0))^{-1}\{p(M-m,m) + \int_0^m q(m-u)\,d_u(1-p(M-m,u))\} \tag{2.1.5}$$

$$E[B] = \int_m^\infty [u-m+b(0)]\,d_u(1-p(M-m,u)) + \int_0^m b(m-u)\,d_u(1-p(M-m,u)). \tag{2.1.6}$$

To derive the relations (2.1.4)-(2.1.6) we conditioned on the undershoot of the inventory level m when this level is downcrossed. In case the undershoot

is less than m the inventory drops to a positive level between 0 and m. Otherwise the inventory on hand is exceeded by the demand of a customer and the inventory level drops to zero.

It remains to find expressions for the basic functions. This will be achieved by the derivation of relations between the basic functions for the present model and those for the backlog model of chapter 1. In the sequel we distinguish between the two sets of basic functions by indexing the latter basic functions by a "B". Once the relations have been established we can invoke the results obtained in the sections 1.3 and 1.4 to obtain the desired expressions for the basic functions for the lost-sales model.

This chapter is further organized as follows. In section 2.2 we will consider the functions $t_1(x)$ and $p(x,u)$. In section 2.3 we focus on the functions $t_2(x)$, $q(x)$ and $b(x)$. In section 2.4 we validate the approximations by computer simulation and study the sensitivity of the switching level m to more than the first two moments of the demand per unit time.

2.2. Expressions for $t_1(x)$ and $p(x,u)$.

We assume that at epoch 0 the inventory level equals x+m, $x \geq 0$ and production rate π_1 is used. Define

τ_1 := the first arrival epoch after epoch 0.

τ_n := the time that elapses between the arrival of the (n-1)-th and n-th customer, $n \geq 2$.

D_n := the demand of the n-th arriving customer.

We again consider the random-walk $\{S_n\}_{n=1}^{\infty}$ defined by

$$S_0 := 0, \quad S_n := \sum_{i=1}^{n} X_i, \quad n \geq 1,$$

where

$$X_n := D_n - \pi_1 \tau_n, \quad n \geq 1.$$

We further recall the definition of $N(x)$,

$$N(x) := \min\{n | S_n > x\}, \quad x \geq 0,$$

i.e. the $N(x)$-th customer is the first one causing an undershoot of level m. It then follows from the redefinition of $p(x,u)$ that

$$(2.2.1) \qquad p(x,u) = P\{S_{N(x)} - x > u\}, \quad x \geq 0.$$

The sequences $\{D_n\}$ and $\{\tau_n\}$ are exogeneous to the inventory process. This implies that the random walk $\{S_n\}$ is not affected by the way excess demand is lost. Thus the process $\{S_n\}$ defined above is identical to the corresponding random walk defined in section 1.3. Therefore we find

$$(2.2.2) \qquad p(x,u) = p_B(x,u), \quad x \geq 0,$$

where $p_B(x,u)$ denotes the counterpart of $p(x,u)$ in the backlog model.

By using the principle of conservation of flow we obtain in exactly the same way as we derived (1.3.10),

$$(2.2.3) \qquad t_1(x) = \frac{x+E[U(x)]}{\lambda E[D]-\pi_1}, \quad x \geq 0,$$

where $E[U(x)]$ is defined by

$$E[U(x)] := \int_0^\infty p(x,u)du.$$

Hence it follows from (1.3.10), (1.3.11) and (2.2.2) that

$$(2.2.4) \qquad t_1(x) = t_{1,B}(x), \quad x \geq 0,$$

where $t_{1,B}(x)$ is the counterpart of $t_1(x)$ for the backlog model.

Now we are in a position to give approximations for $t_1(M-m)$ and $p(M-m,u)$, $u \geq 0$. We distinguish between the cases M=m and M>m.

Case M=m:

It follows from (1.3.33), (1.3.34), (2.2.2) and (2.2.4) that

$$(2.2.5)\qquad t_1(0) = \begin{cases} 0 & \text{when } \pi_1 < 0 \\ 1/\lambda & \text{when } \pi_1 = 0, \\ 1/(\pi_1 s^*) & \text{when } \pi_1 > 0 \end{cases}$$

and, for any $u \geq 0$,

$$(2.2.6)\qquad p(0,u) = \begin{cases} 0 & \text{when } \pi_1 < 0 \\ 1-F(u) & \text{when } \pi_1 = 0. \\ \lambda/\pi_1 \int_0^\infty e^{-s^* y}(1-F(y+u))dy & \text{when } \pi_1 > 0 \end{cases}$$

Case M>m:

To use the approximations 1.3.1 and 1.3.2 we have to restrict ourselves to (m,M)-policies satisfying

Condition 2.2.1.

For the case of $\pi_1 \leq 0$

$$M-m \geq \begin{cases} E[D]-\pi_1/\lambda & \text{when } c_D^2 \leq 1 \\ 1\tfrac{1}{2}c_D^2(E[D]-\pi_1/\lambda) & \text{when } c_D^2 > 1 \end{cases}$$

and for the case of $\pi_1 > 0$

$$M-m \geq \begin{cases} E[Z_1] & \text{when } c_{Z_1}^2 \leq 1 \\ 1\tfrac{1}{2}c_{Z_1}^2 E[Z_1] & \text{when } c_{Z_1}^2 > 1 \end{cases}$$

where $c_{Z_1}^2 = (E[Z_1^2]-(E[Z_1])^2)/(E[Z_1])^2$ and $E[Z_1]$ and $E[Z_1^2]$ are given by (1.3.22)[1] and (1.3.23), respectively. Then it follows from approximations 1.3.1 and 1.3.2 and the equations (2.2.2) and (2.2.4) that

Approximation 2.2.1.

$$t_1(M-m) \cong \begin{cases} \dfrac{M-m}{\lambda E[D]-\pi_1} + \dfrac{\lambda E[D^2]}{2(\lambda E[D]-\pi_1)^2} & \text{when } \pi_1 \leq 0 \\ \dfrac{M-m}{\lambda E[D]-\pi_1} + \dfrac{\lambda E[D^2]}{2(\lambda E[D]-\pi_1)^2} - \dfrac{1}{s^*(\lambda E[D]-\pi_1)} & \text{when } \pi_1 > 0 \end{cases}$$

Approximation 2.2.2. For any $u \geq 0$

$$p(M-m,u) \cong \begin{cases} \frac{\lambda}{\lambda E[D]-\pi_1} \int_u^\infty (1-F(y))dy & \text{when } \pi_1 \leq 0 \\ \frac{\lambda}{\lambda E[D]-\pi_1} \int_u^\infty (1-F(y))(1-e^{-s^*(y-u)})dy, & \text{when } \pi_1 > 0 \end{cases}$$

The constant s^* is the unique positive root of

$$(2.2.7) \qquad s - \frac{\lambda}{\pi_1}(1 - \int_0^\infty e^{-sy}dF(y)) = 0.$$

2.3. Expressions for $t_2(x)$, $b(x)$ and $q(x)$.

Throughout this section we assume that production rate π_2 is used and at epoch 0 the inventory level equals $0 \leq x \leq M$. We must find expressions for $t_2(x)$, $b(x)$ and $q(x)$. We first derive relations between these functions and the corresponding basic functions $t_{2,B}(x)$, $b_B(x)$ and $q_B(x)$ in the backlog model. The functions $t_{2,B}(x)$ and $q_B(x)$ are defined analogously to $t_2(x)$ and $q(x)$, while the function $b_B(x)$ is defined by

$b_B(x)$:= the expected amount of demand that is backlogged until the inventory level reaches M, $\quad 0 \leq x \leq M$.

In section 1.4 we derived tractable expressions for $t_{2,B}(x)$, $q_B(x)$ and $b_B(x)$.

First we turn our attention to the hitting probability $q(x)$. Once we have fixed an (m,M)-policy the evaluation of the inventory process is completely determined by the sequences $\{D_n\}$ and $\{\tau_n\}$. Given a positive initial inventory and a realization of the sequences $\{D_n\}$ and $\{\tau_n\}$ it then follows that the induced sample paths for the backlog model and the lost-sales model are identical *as long as the inventory is positive*. Hence, given a realization of the sequences $\{D_n\}$ and $\{\tau_n\}$ and a positive initial inventory $x \leq M$ we have the following result:

> *A stockout occurs before the inventory level reaches the value M in the lost-sales model* if and only if *a stockout occurs before the inventory level reaches the value M in the backlog model.*

This implies that for all $x \geq 0$,

$$(2.3.1) \qquad q(x) = q_B(x).$$

The equations (1.2.11), (2.1.5), (2.2.2) and (2.3.1) together imply that

$$(2.3.2) \qquad E[Q] = E[Q_B].$$

Here Q_B denotes the number of stockout occurences in a cycle for the backlog model.

To derive approximations for $t_2(x)$ and $b(x)$ we derive an exact relation between these two functions by using the principle of conservation of flow. Recall the definition of $T_2(x)$,

$T_2(x)$:= the time until the inventory level reaches the value M, $\quad 0 \leq x \leq M$.

Then, by its very definition,

$$(2.3.3) \qquad t_2(x) = E[T_2(x)].$$

Clearly, the production in $(0,T_2(x)]$ is equal to $\pi_2 T_2(x)$. The demand in $(0,T_2(x)]$ is equal to $V(T_2(x))$ and the amount of demand that is lost in this time interval is equal to $B(T_2(x))$. Now, by the priniple of conservation of flow, we have that at any time $t>0$ the inventory position $X(t)$ equals the sum of the initial inventory $X(0)$ and the amount produced in $(0,t]$ minus *the amount of demand that is not lost* during $(0,t]$. Using $X(0)=x$ and $X(T_2(x))=M$ this implies that

$$(2.3.4) \qquad M = x + \pi_2 T_2(x) - (V(T_2(x)) - B(T_2(x))), \quad 0 \leq x \leq M.$$

We also have that

$$(2.3.5) \qquad b(x) = E[B(T_2(x))], \quad 0 \leq x \leq M$$

and

$$(2.3.6) \qquad V(T_2(x)) = \sum_{n=1}^{N(T_2(x))} D_n, \quad 0 \leq x \leq M.$$

Though $N(T_2(x))$ is not a stopping time for the sequence $\{\tau_n\}$ it can be verified that $N(T_2(x))$ is indeed a stopping time for $\{D_n\}$. Application of Wald's equation yields

$$E[V(T_2(x))] = E[N(T_2(x))].E[D], \quad 0 \leq x \leq M. \tag{2.3.7}$$

The Markov property of the exponential distribution implies that after the inventory level has reached the value M an exponentially distributed time with mean $1/\lambda$ elapses until the next customer arrives. Hence

$$E\Big[\sum_{n=1}^{N(T_2(x))+1} \tau_n - T_2(x)\Big] = \frac{1}{\lambda}, \quad 0 \leq x \leq M. \tag{2.3.8}$$

The fact that $N(T_2(x))+1$ is a stopping time for $\{\tau_n\}$ allows for the application of Wald's equation, yielding

$$E[N(T_2(x))] = \lambda t_2(x), \tag{2.3.9}$$

where we used (2.3.3) and (2.3.8). Combining the equations (2.3.7) and (2.3.9) we obtain

$$E[V(T_2(x))] = \lambda E[D] t_2(x). \tag{2.3.10}$$

Taking expectations in (2.3.4) and using (2.3.3), (2.3.5) and (2.3.10) we find

$$M = x + \pi_2 t_2(x) - (\lambda E[D] t_2(x) - b(x)),$$

which implies

$$t_2(x) = \frac{M-x-b(x)}{\pi_2 - \lambda E[D]}. \tag{2.3.11}$$

To find another relation for $t_2(x)$ we will express $t_2(x)$ in terms of basic functions for the backlog model. Let us define for the backlog model

$\bar{t}_B(x)$:= the expected time during which the inventory is negative until the inventory reaches the level M for the first time, $0 \leq x \leq M$.

The following result holds,

$$(2.3.12) \qquad t_{2,B}(x) = t_2(x) + t_B^-(x).$$

We shall give a rough outline of a proof of this result.

We first define for the lost-sales model

$Q(x)$:= the number of stockouts until the inventory level reaches the value M, when the initial inventory is x, $0 \leq x \leq M$.

Also, for the backlog model let

$Q_B(x)$:= the number of stockouts until the inventory level reaches the value M, when the initial inventory is x, $0 \leq x \leq M$.

By the same arguments as used to derive (1.2.11) and using (2.3.1) we find

$$(2.3.13) \qquad Q(x) \overset{d}{=} Q_B(x), \quad 0 \leq x \leq M,$$

where the notation $X \overset{d}{=} Y$ means that $P\{X \leq z\} = P\{Y \leq z\}$, for all z.

Next we define for the lost-sales model the random variables σ_i and ν_i,

$$\nu_0 := 0, \quad \sigma_i := \inf\{t : t > \nu_{i-1},\ X(t) \leq 0\}, \quad i \geq 1$$

$$\nu_i := \inf\{t : t \geq \sigma_i,\ X(t) \geq 0\}, \quad i \geq 1.$$

Similarly, we define the random variables $\sigma_{B,i}$ and $\nu_{B,i}$ for the backlog model. In words σ_i is the i^{th} epoch at which a stockout occurs, ν_i is the first epoch beyond the i^{th} stockout at which the inventory level equals 0. For ease of notation we suppress the dependency on $X(0)=x$. For the lost-sales model we have

$$(2.3.14) \qquad \nu_i = \sigma_i, \quad i \geq 1, \text{ with probability 1.}$$

Then, using the above definitions, it follows after some reflections that for the backlog model,

$$(2.3.15)\qquad t_B^-(x) = E[\sum_{i=1}^{Q_B(x)} (\nu_{B,i}-\sigma_{B,i})], \quad 0 \leq x \leq M$$

and

$$(2.3.16)\qquad t_2^B(x) = t_B^-(x) + E[\sum_{i=1}^{Q_B(x)} (\sigma_{B,i}-\nu_{B,i-1})+T_{2,B}(x)-\nu_{B,Q_B(x)}], \; 0 \leq x \leq M,$$

where $T_{2,B}(x)$ is defined analogously to $T_2(x)$. For the lost-sales model we find

$$(2.3.17)\qquad t_2(x) = E[\sum_{i=1}^{Q(x)} (\sigma_i-\nu_{i-1}) + T_2(x) - \nu_{Q(x)}], \quad 0 \leq x \leq M,$$

where we used relation (2.1.14).

Using the lack of memory of the Poisson arrival process it can be shown that the respective inventory processes $\{X(t),\ t\geq 0\}$ and $\{X_B(t),\ t\geq 0\}$ for the lost-sales and backlog model behave probabilistically the same *if we only observe the process* $\{X_B(t),\ t\geq 0\}$ *at times when the inventory is non-negative*. Then it can be shown that

$$(2.3.18)\qquad \sigma_1 \overset{d}{=} \sigma_{B,1}$$

$$(2.3.19)\qquad (\sigma_i-\nu_{i-1})1_{\{Q\geq i\}} \overset{d}{=} (\sigma_{B,i}-\nu_{B,i})1_{\{Q_B\geq i\}}$$

$$(2.3.20)\qquad T_2(x) - \nu_{Q(x)} \overset{d}{=} T_{2,B}(x) - \nu_{B,Q_B(x)}.$$

Combining (2.3.13), (2.3.15)-(2.3.20) we finally obtain equation (2.3.12).

In section 1.2 we argued that $E[B]=\pi_2 E[J]$. Using the same arguments we obtain

$$(2.3.21)\qquad b_B(x) = \pi_2 t_B^-(x), \quad 0 \leq x \leq M.$$

Note that equation (2.3.21) does not hold for $x<0$ because of the initial backlog. The equations (1.4.1), (2.3.12) and (2.3.21) together imply

$$(2.3.22)\qquad t_2(x) = \frac{M-x}{\pi_2-\lambda E[D]} - \frac{b_B(x)}{\pi_2}, \quad 0 \leq x \leq M.$$

An expression for b(x) in terms of $b_B(x)$ can be deduced from (2.3.11) and (2.3.22),

$$(2.3.23) \qquad b(x) = \frac{\pi_2-\lambda E[D]}{\pi_2} \cdot b_B(x), \quad 0 \leq x \leq M.$$

Now we are in a position to give approximations for q(x), b(x) and $t_2(x)$. In view of relations (2.3.1), (2.3.22) and (2.3.23) we want to use approximations 1.4.1 and 1.4.2. Therefore we must assume that the distribution assumption DA holds, i.e.

$$(2.3.24) \qquad 1-F(x) = O(e^{-\kappa x}), \ (x\to\infty), \quad \text{for some } \kappa > 0.$$

Under this assumption we can define the number δ as the unique positive solution to the equation

$$(2.3.25) \qquad \int_0^\infty e^{\delta y} \frac{\lambda}{\pi_2} (1-F(y))dy = 1$$

Also, because of (2.3.24), the number ν defined by

$$(2.3.26) \qquad \nu := \int_0^\infty y e^{\delta y} \frac{\lambda}{\pi_2} (1-F(y))dy$$

is finite.
Thus we find

Approximation 2.3.1. For all $0 \leq x \leq M$,

$$q(x) = \frac{q_\infty(x)-q_\infty(M)}{1-q_\infty(M)}$$

with

$$q_\infty(x) \cong \alpha e^{-\beta x} + \frac{(\pi_2-\lambda E[D])}{\pi_2 \delta \nu} e^{-\delta x}$$

and

$$(2.3.27) \qquad \alpha = \frac{\lambda E[D]}{\pi_2} - \frac{(\pi_2-\lambda E[D])}{\pi_2 \delta \nu},$$

$$\beta = \left[\frac{\lambda E[D]}{\pi_2} - \frac{(\pi_2-\lambda E[D])}{\pi_2 \delta \nu}\right]\left[\frac{\lambda E[D^2]}{2(\pi_2-\lambda E[D])} - \frac{(\pi_2-\lambda E[D])}{\pi_2 \delta^2 \nu}\right]^{-1}.$$

Approximation 2.3.2. For all $0 \leq x \leq M$,

$$b(x) = b_\infty(x) - b_\infty(M)$$

with

$$(2.3.28) \qquad b_\infty(x) \cong \alpha/\beta e^{-\beta x} + \frac{\pi_2 - \lambda E[D]}{\pi_2 \delta^2 \nu} e^{-\delta x}$$

and the constants α and β are given by (2.3.27).

Approximation 2.3.3. For all $0 \leq x \leq M$

$$t_2(x) = \frac{M-x}{\pi_2 - \lambda E[D]} - \frac{(b_\infty(x) - b_\infty(M)}{\pi_2 - \lambda E[D]}$$

and $b_\infty(x)$ is approximated by the right-hand side of (2.3.28)

In the next section we present numerical results, showing the accuracy of the approximations. We also consider the sensitivity of the service level m to more than the first two moments of the demand per unit time.

2.4. Numerical results and conclusions.

In this chapter we analysed the lost-sales model by establishing first exact relations between this model and the backlog model and by invoking next the approximations derived in chapter 1 for the backlog model. In view of this analysis we may expect that the resulting approximations for the lost-sales model show a similar performance of a good quality as the approximations for the backlog model. Our numerical investigations indeed support this claim and also show that concerning insensitivity issues the same conclusions can be made as for the backlog model.

In the tables 2.4.1 and 2.4.2 we give numerical results showing the accuracy of the approximations in the lost-sales model for both the β-service measure and the γ-service measure. The parameters of the model are varied as follows. The production rate π_1 has the three values -0.5, 0 and 0.5, the production rate π_2 has the three values 1.25, 2 and 5. The service levels β and γ are varied as 0.95 and 0.99. In all examples we have chosen λ=1 and E[D]=1. As before c_D denotes the coefficient of variation of the demand size D. We vary c_D^2 as 0, 1/3, 2/3 and 2 and consider the following demand distributions,

(i) deterministic demand ($c_D^2=0$).
(ii) gamma demand ($c_D^2=1/3$, 2/3 and 2).

The value of M-m is predetermined from the formula (1.5.1) with K=25 and h=1. Next, by applying the approximations given in sections 2.2 and 2.3, we compute the switching level m such that the (m,M)-rule meets the β-service level requirement or the γ-service level requirement. As in section 1.5 we used the exact expression for p(M-m,u) and $t_1(M-m)$ when the demand D is deterministic. For the approximate (m,M)-rules the actual service levels β_{act} and γ_{act} were obtained by computer simulation, where in each example 250,000 customer demands were simulated. As before, the notation 0.950(4) for the actual service level means that the 95% confidence interval of the simulated value is given by 0.946-0.954.

It is noteworthy from the tables 2.4.1 and 2.4.2 and the tables 1.5.1 and 1.5.2 that the values of the switching level m in the lost-sales model differ significantly from the corresponding values in the backlog model. This phenomenon is peculiar to the production-inventory model with continuously occurring inventory replenishments; for the pure inventory model with discrete replenishments the lost-sales model and the backlog model are nearly the same when the required service level is high, see Tijms and Groenevelt [1984].

Finally we make the following remarks on the results in the tables 2.4.1 and 2.4.2. In table 2.4.1 the case of $c_D^2=0$, $\pi_1=0.5$, $\pi_2=5$ and $\beta=0.95$ is marked with an asterisk. In this particular case it turned out that the (m,M)-rules with M-m predetermined by (1.5.1) provide a service level $\beta>0.95$ for all $m\geq 0$. The approximate β-service level of the (0.00,4.71)-rule is 0.956. In table 2.4.2 the case of $c_D^2=0$, $\pi_1=0$, $\pi_2=5$ and $\gamma=0.95$ is marked with a double asterisk. For this case the prespecified level $\gamma=0.95$ cannot be achieved. This results from the fact that for deterministic demand and $\pi_1=0$ the fraction of customers whose demand is met directly from stock on hand is a discontinuous function of m when M-m is fixed. The approximate γ-service level of the (0.68,7.00)-rule is 0.971.

Table 2.4.1. The approximate (m,M)-rules and their actual β-service levels.

			$c_D^2=0$			$c_D^2=1/3$		
π_1	π_2	β	m	M	β_{act}	m	M	β_{act}
-0.5	1.25	0.95	2.26	5.53	0.949(2)	3.51	6.78	0.951(2)
-0.5	2	0.95	0.52	5.99	0.950(1)	1.06	6.53	0.950(2)
-0.5	5	0.95	0.16	7.54	0.950(1)	0.42	7.81	0.951(1)
0	1.25	0.95	2.43	5.59	0.951(2)	3.46	6.63	0.950(2)
0	2	0.95	0.93	5.93	0.950(2)	1.11	6.11	0.950(2)
0	5	0.95	0.37	6.69	0.950(1)	0.50	6.83	0.951(2)
0.5	1.25	0.95	1.83	4.72	0.950(2)	2.94	5.82	0.950(2)
0.5	2	0.95	0.26	4.34	0.949(1)	0.74	4.83	0.951(2)
0.5	5	0.95	0.00	4.71	0.956(1)*	0.18	4.90	0.951(2)
-0.5	1.25	0.99	5.66	8.93	0.990(1)	8.22	11.49	0.991(2)
-0.5	2	0.99	1.76	7.24	0.990(1)	2.93	8.40	0.990(1)
-0.5	5	0.99	0.79	8.17	0.990(1)	1.51	8.89	0.990(1)
0	1.25	0.99	5.83	8.99	0.989(2)	8.17	11.34	0.991(2)
0	2	0.99	2.18	7.18	0.990(1)	2.98	7.98	0.990(1)
0	5	0.99	0.86	7.19	0.990(1)	1.58	7.90	0.990(1)
0.5	1.25	0.99	5.24	8.12	0.990(2)	7.65	10.53	0.991(2)
0.5	2	0.99	1.51	5.59	0.990(1)	2.61	6.70	0.990(1)
0.5	5	0.99	0.58	5.30	0.989(1)	1.26	5.98	0.990(1)

			$c_D^2=2/3$			$c_D^2=2$		
π_1	π_2	β	m	M	β_{act}	m	M	β_{act}
-0.5	1.25	0.95	4.79	8.06	0.951(3)	10.05	13.32	0.951(3)
-0.5	2	0.95	1.69	7.17	0.950(2)	4.63	10.11	0.950(3)
-0.5	5	0.95	0.76	8.14	0.949(2)	2.57	9.96	0.948(3)
0	1.25	0.95	4.72	7.88	0.949(2)	9.88	13.04	0.949(4)
0	2	0.95	1.74	6.74	0.949(2)	4.63	9.63	0.950(3)
0	5	0.95	0.86	7.18	0.950(2)	2.72	9.05	0.949(2)
0.5	1.25	0.95	4.07	6.95	0.951(3)	8.71	11.59	0.953(3)
0.5	2	0.95	1.30	5.38	0.950(2)	3.80	7.89	0.950(2)
0.5	5	0.95	0.49	5.20	0.950(2)	2.09	6.81	0.949(3)
-0.5	1.25	0.99	10.81	14.08	0.990(2)	21.32	24.59	0.991(2)
-0.5	2	0.99	4.19	9.67	0.990(1)	9.70	15.18	0.990(2)
-0.5	5	0.99	2.30	9.68	0.990(1)	5.96	13.34	0.989(2)
0	1.25	0.99	10.74	13.90	0.990(2)	21.15	24.32	0.990(2)
0	2	0.99	4.24	9.24	0.990(1)	9.71	14.71	0.990(2)
0	5	0.99	2.40	8.72	0.990(1)	6.11	12.43	0.991(2)
0.5	1.25	0.99	10.09	12.97	0.991(2)	19.98	22.87	0.989(2)
0.5	2	0.99	3.80	7.88	0.989(1)	8.87	12.96	0.989(2)
0.5	5	0.99	2.02	6.74	0.990(1)	5.48	10.20	0.990(2)

Table 2.4.2. The approximate (m,M)-rules and their actual γ-service levels.

			$c_D^2=0$			$c_D^2=1/3$		
π_1	π_2		m	M	γ_{act}	m	M	γ_{act}
-0.5	1.25	0.95	3.82	7.10	0.949(3)	4.73	8.01	0.951(2)
-0.5	2	0.95	1.21	6.68	0.950(2)	1.65	7.13	0.950(2)
-0.5	5	0.95	0.71	8.09	0.950(2)	0.84	8.23	0.950(2)
0	1.25	0.95	3.99	7.15	0.951(3)	4.69	7.85	0.950(2)
0	2	0.95	1.64	6.64	0.946(2)	1.70	6.70	0.950(2)
0	5	0.95	0.68	7.00	0.968(2)**	0.93	7.25	0.950(2)
0.5	1.25	0.95	3.40	6.28	0.951(3)	4.16	7.05	0.951(2)
0.5	2	0.95	0.94	5.02	0.948(2)	1.34	5.42	0.951(2)
0.5	5	0.95	0.47	5.18	0.951(1)	0.60	5.32	0.951(2)
-0.5	1.25	0.99	7.39	10.67	0.990(2)	9.60	12.87	0.990(2)
-0.5	2	0.99	2.49	7.97	0.990(1)	3.54	9.02	0.990(1)
-0.5	5	0.99	1.21	8.60	0.990(1)	1.93	9.32	0.989(1)
0	1.25	0.99	7.56	10.72	0.991(2)	9.55	12.71	0.991(1)
0	2	0.99	2.91	7.91	0.990(1)	3.59	8.59	0.990(1)
0	5	0.99	1.34	7.66	0.988(1)	2.01	8.33	0.990(1)
0.5	1.25	0.99	6.97	9.86	0.991(2)	9.02	11.91	0.991(2)
0.5	2	0.99	2.24	6.32	0.990(1)	3.23	7.31	0.990(1)
0.5	5	0.99	0.98	5.69	0.989(1)	1.69	6.40	0.990(1)

			$c_D^2=2/3$			$c_D^2=2$		
π_1	π_2		m	M	γ_{act}	m	M	γ_{act}
-0.5	1.25	0.95	5.47	8.75	0.951(3)	7.40	10.67	0.951(2)
-0.5	2	0.95	2.03	7.51	0.950(2)	3.22	8.70	0.950(2)
-0.5	5	0.95	1.00	8.38	0.950(2)	1.58	8.96	0.950(2)
0	1.25	0.95	5.40	8.57	0.949(2)	7.23	10.39	0.949(3)
0	2	0.95	2.08	7.08	0.949(2)	3.22	8.22	0.950(2)
0	5	0.95	1.10	7.43	0.950(2)	1.71	8.04	0.950(2)
0.5	1.25	0.95	4.75	7.64	0.952(3)	6.06	8.94	0.952(2)
0.5	2	0.95	1.64	5.73	0.950(2)	2.40	6.48	0.950(2)
0.5	5	0.95	0.73	5.44	0.950(1)	1.07	5.79	0.950(2)
-0.5	1.25	0.99	11.59	14.86	0.990(2)	18.17	21.44	0.991(1)
-0.5	2	0.99	4.55	10.03	0.990(1)	8.17	13.65	0.990(2)
-0.5	5	0.99	2.56	9.94	0.990(1)	4.83	12.21	0.990(1)
0	1.25	0.99	11.52	14.68	0.990(2)	18.00	21.17	0.990(2)
0	2	0.99	4.60	9.60	0.990(1)	8.17	13.17	0.990(1)
0	5	0.99	2.66	8.98	0.990(1)	4.97	11.30	0.990(1)
0.5	1.25	0.99	10.87	13.75	0.991(2)	16.83	19.72	0.990(2)
0.5	2	0.99	4.16	8.24	0.989(1)	7.34	11.42	0.989(2)
0.5	5	0.99	2.28	7.00	0.990(1)	4.35	9.06	0.990(1)

Our numerical results indicate that, concerning the sensitivy of the switching level m to more than the first two moments of the customer demand D, the same conclusions can be made as for the backlog model (see section 1.5). A different, but related sensitivity study concerns the influence of the arrival rate λ on the switching level m while fixing the first two moments of the demand per unit time. Such a sensitivity analysis may be important in situations in which only limited information is available about the demand process, e.g. we know only the first two moments of the demand during a given time interval (0,t].

Assuming that the process $\{V(t),\ t\geq 0\}$ with

$$V(t) := \text{the total demand in } (0,t]$$

is a compound Poisson process with arrival rate λ and individual demand size D, it follows that

$$(2.4.1) \qquad E[V(t)] = \lambda E[D]t, \quad E[V^2(t)] = \lambda E[D^2]t + (\lambda E[D])^2 t^2.$$

Then $E[V(t)]$ and $E[V^2(t)]$ are completely determined by $\lambda E[D]$ and $\lambda E[D^2]$.

Let us define

$$\mathcal{V} := V(1),$$

i.e. $\mathcal{V}$ is the total demand during one unit of time. From (2.4.1) we obtain

$$(2.4.2) \qquad E[\mathcal{V}] = \lambda E[D], \quad c_{\mathcal{V}}^2 = \lambda^{-1}(c_D^2+1).$$

Note that $\lambda \geq (c_{\mathcal{V}}^2)^{-1}$. Suppose we have estimated $E[V(t)]$ and $E[V^2(t)]$. Then we can determine $E[\mathcal{V}]$ and $c_{\mathcal{V}}^2$. The question arises whether this information is sufficient to determine an (m,M)-rule that satisfies a given β-service level constraint. To answer this question we fix the values of $E[\mathcal{V}]$ and $c_{\mathcal{V}}^2$ and we determine the switch-over level m for several values of the arrival rate λ. In all examples we have chosen $E[\mathcal{V}]=1$. The production rate π_1 has the two values 0 and 0.5, the production rate π_2 has the three values 1.25, 2 and 5. The service level β is varied as 0.95 and 0.99. We vary $c_{\mathcal{V}}^2$ as 1, 2, 4 and 8. For each combination of these model parameters we vary λ as 1, 2 and 4. For given values of λ, $E[\mathcal{V}]$ and $c_{\mathcal{V}}^2$ we compute from (2.4.2) the values of E[D] and c_D^2. We fit to the demand D a

deterministic distribution when $c_D^2=0$ and a gamma distribution when $c_D^2>0$. Then we compute the switch-over level m that satisfies the given β-service level constraint.

Table 2.4.3. Sensitivity of m to the first two moments of the demand per unit time.

			$c_V^2=1$			$c_V^2=2$		
π_1	π_2	β	λ=1	λ=2	λ=4	λ=1	λ=2	λ=4
0	1.25	0.95	2.43	2.33	2.37	5.99	6.10	6.15
0	2	0.95	0.93	0.64	0.68	2.42	2.57	2.64
0	5	0.95	0.37	0.25	0.26	1.27	1.40	1.45
0.5	1.25	0.95	1.83	1.98	2.05	5.21	5.37	5.45
0.5	2	0.95	0.26	0.42	0.49	1.89	2.10	2.19
0.5	5	0.95	0.00	0.05	0.10	0.84	1.04	1.13
0	1.25	0.99	5.83	5.99	6.17	13.32	13.71	13.90
0	2	0.99	2.18	2.21	2.41	5.56	6.07	6.31
0	5	0.99	0.86	1.25	1.41	3.27	3.75	3.95
0.5	1.25	0.99	5.24	5.65	5.85	12.54	12.98	13.20
0.5	2	0.99	1.51	1.99	2.22	5.03	5.59	5.85
0.5	5	0.99	0.58	1.05	1.26	2.84	3.40	3.66
			$c_V^2=4$			$c_V^2=8$		
π_1	π_2	β	λ=1	λ=2	λ=4	λ=1	λ=2	λ=4
0	1.25	0.95	13.84	13.97	14.04	30.01	30.18	30.26
0	2	0.95	6.99	7.18	7.27	17.00	17.24	17.36
0	5	0.95	4.37	4.58	4.68	11.90	12.19	12.34
0.5	1.25	0.95	12.27	12.45	12.55	26.82	27.04	27.15
0.5	2	0.95	5.83	6.08	6.19	14.46	14.74	14.89
0.5	5	0.95	3.51	3.79	3.92	9.92	10.27	10.44
0	1.25	0.99	29.06	29.48	29.69	61.02	61.48	61.72
0	2	0.99	13.99	14.55	14.82	31.78	32.40	32.70
0	5	0.99	9.14	9.67	9.91	22.20	22.80	23.09
0.5	1.25	0.99	27.49	27.96	28.20	57.84	58.35	58.61
0.5	2	0.99	12.84	13.44	13.73	29.22	29.86	30.18
0.5	5	0.99	8.29	8.91	9.20	20.26	20.94	21.28

The results displayed in table 2.4.3 are quite surprising. We observe that as c_V^2 increases the switching level m increases, but the differences between the three switching levels for the cases of λ=1, 2 and 4 increase only very slightly. Exceptions to this finding are the cases of $\pi_1=0$, β=0.95 and $\pi_2=2$ and 5, where the difference between the m-values for λ=1

and 2 decreases as c_V^2 increases from 1 to 2. This is due to the fact that for the case of $\lambda=1$ and $c_V^2=1$ the demand D is deterministic. The above insensitivity result is surprising, since the values of c_D^2 corresponding to $\lambda=1$, 2 and 4 differ significantly and the switching level m is known to be rather sensitive to more than the first two moments of the demand D when c_D^2 gets large. An intuitive explanation for the insensitivity result is as follows. Keeping $E[V]$ and c_V^2 fixed, an increase of λ causes both the decrease of $E[D]$ and an increase of c_D^2. These opposed effects neutralize each other in the following sense: The expected undershoot and the expected lost demand per dissatisfied customer remain approximately the same. Also, when M is sufficiently large the number of stockouts per cycle remains approximately the same.

We conclude that, knowing the type of demand distribution, the switch-over level m becomes less sensitive to λ as c_V^2 increases with $E[V]$ kept fixed under a given β-service level constraint.

In practical situations, where we have estimates only for $E[V]$ and c_V^2, it is unlikely that we can specify the type of demand distribution. In view of the sensitivity results given in section 1.5.1 the type of demand distribution is irrelevant when the demand is non-erratic i.e. $c_D^2 \leq 1$. Thus we see that when $c_D^2 \leq 1$ and c_V^2 large the switching level m is insensitive to the value of λ. However, for the case of $c_D^2 \leq 1$ it follows from (2.4.2) that

$$(2.4.3) \qquad (c_V^2)^{-1} \leq \lambda \leq 2(c_V^2)^{-1}$$

and hence for the case of $c_D^2 \leq 1$ and c_V^2 large we have that λ is small and, moreover, (2.4.3) provides a rather narrow range for λ.

Concluding, in situations where only estimates of $E[V]$ and c_V^2 are known, an (m,M)-rule satisfying a given β-service level constraint can be given without having an estimate of the arrival rate λ when the demand is non-erratic and the arrival rate is small. This corresponds to the case of slow moving items for which the demand is non-erratic. In all other cases it is necessary to estimate λ. The same conclusions hold with respect to the α-service level. However, with respect to the γ-service level the switching level m is very sensitive to λ when keeping $E[V]$ and c_V^2 fixed under a given γ-service level constraint.

3. THE SINGLE PRODUCT PRODUCTION-INVENTORY MODEL WITH MIXED BACKLOGGING AND PARTIAL LOST-SALES.

In this chapter we consider the single product production-inventory model in which the backlog at any time must not exceed a given amount L_0. If a customer arrives by whose demand the backlog would exceed this critical level, then the inventory drops just to the level $-L_0$ and that part of the demand by which the level $-L_0$ would be undershot is lost. The backlog and lost-sales model discussed in the chapters 1 and 2 are special cases of this model with $L_0=\infty$ and $L_0=0$, respectively. We will show below that the service measures associated with this model can be expressed in terms of the basic functions that determine the service measures in the backlog and lost-sales model. Once these relations have been established, the approximations given in the chapters 1 and 2 can be used to find tractable expressions for the service measures of the more general model considered in this section.

3.1. Model and service measures.

We consider a single product production-inventory system in which the customers arrive according to a Poisson process with rate λ. The demands of arriving customers are independent random variables with common distribution function F. The demand sizes are independent of the arrival process.

The production facility continually adds the product to the inventory by using one out of two production rates π_1 and π_2 such that

$$(3.1.1) \qquad \pi_1 < \lambda E[D] < \pi_2.$$

The generic random variable D denotes the demand of a single customer. The inventory is controlled by an (m,M)-rule with $0 \leq m \leq M$. The system has an infinite storage capacity. The maximum backlog at an arbitrary point in time must not exceed a given amount L_0.

To describe precisely the way excess demand is handled, consider a customer arriving when the inventory equals x. Let D denote the demand of this customer. The inventory is decreased by an amount of $\min(D,x+L_0)$ and an amount of $\max(0,D-x-L_0)$ of the excess demand is lost.

As in the lost-sales model we note that the inequality $\pi_2 > \lambda E[D]$ in (3.1.1) is not necessary, since the inventory level cannot drift to $-\infty$ because of the boundary $-L_0$. However, in case of a service measure, one typically needs that $\pi_2 > \lambda E[D]$ indeed holds when a sufficiently high service level is required.

This model was analysed in the paper by Doshi et al [1978] for both finite and infinite storage capacity (for the finite capacity case see remark 7.5.2). Using a renewal-theoretic approach expressions were given for the long-run average costs. However, these expressions lead to tractable results only for the special case of exponentially distributed demand.

Fix an (m,M)-rule with $0 \leq m \leq M$. Define for $t \geq 0$.

$N(t)$:= the number of customers arriving in $(0,t]$.

$V(t)$:= the total demand in $(0,t]$.

$X(t)$:= the inventory level at time t.

$B(t)$:= the amount of demand in $(0,t]$ that cannot be met directly from stock on hand.

$B_1(t)$:= the amount of excess demand in $(0,t]$ that is backlogged.

$B_2(t)$:= the amount of excess demand in $(0,t]$ that is lost.

$S(t)$:= the number of customers arriving in $(0,t]$, whose demands cannot be met directly from stock on hand.

$S_1(t)$:= the number of customers arriving in $(0,t]$, whose demands cannot be met directly from stock on hand, but are ultimately satisfied completely.

$S_2(t)$:= the number of customers arriving in $(0,t]$, whose demands are not ultimately satisfied.

$Q(t)$:= the number of stockout occurrences in $(0,t]$.

$C(t) := -\int_0^t X(s)1_{\{X(s)<0\}}ds$ = the cumulative backlog at time t.

$J(t)$:= the amount of time in $(0,t]$, during which the inventory is negative.

We first note that the backlog at an arbitrary point in time never exceeds L_0,

(3.1.2) $\quad X(t) \geq -L_0, \quad \forall t \geq 0.$

Obviously,

(3.1.3) $\quad B(t) = B_1(t) + B_2(t), \quad S(t) = S_1(t) + S_2(t).$

We define the following service measures.

(i) <u>α-service measure</u>.
the long-run average number of stockouts per unit time,

$$\lim_{t\to\infty} \frac{Q(t)}{t}.$$

(ii) <u>β-service measure</u>.
the long-run fraction of demand that cannot be met directly from stock on hand,

$$\lim_{t\to\infty} \frac{B(t)}{V(t)}.$$

(iii) <u>β_1-service measure</u>.
the long-run fraction of demand that is backlogged,

$$\lim_{t\to\infty} \frac{B_1(t)}{V(t)}.$$

(iv) β_2-service measure.
the long-run fraction of demand that is lost,

$$\lim_{t\to\infty} \frac{B_2(t)}{V(t)} .$$

(v) γ-service measure.
the long-run fraction of customers whose demands cannot be met directly from stock on hand,

$$\lim_{t\to\infty} \frac{S(t)}{N(t)} .$$

(vi) γ_1-service measure.
the long-run fraction of customers whose demands cannot be met directly from stock on hand, but are ultimately satisfied completely,

$$\lim_{t\to\infty} \frac{S_1(t)}{N(t)} .$$

(vii) γ_2-service measure.
the long-run fraction of customers whose demands are not ultimately satisfied,

$$\lim_{t\to\infty} \frac{S_2(t)}{N(t)} .$$

(viii) δ-service measure.
the long-run average backlog at an arbitrary point in time,

$$\lim_{t\to\infty} \frac{C(t)}{t} .$$

From the point of view of the customers all service measures are of interest. However, from the point of view of the production facility especially the β_2- and γ_2-service measures are crucial, since these measures reflect the real losses of the system.

As in chapter 1 and 2 we define

a cycle := the time that elapses between two consecutive epochs at which the production rate is switched from π_2 to π_1.

Assuming that at epoch 0 such a regeneration cycle starts, define

T := the time until the production rate is switched from π_2 to π_1,

$N := N(T)$, $V := V(T)$, $B := B(T)$, $B_1 := B_1(T)$, $B_2 := B_2(T)$,

$S := S(T)$, $S_1 := S_1(T)$, $S_2 := S_2(T)$, $Q := Q(T)$, $J := J(T)$,

$C := C(T)$.

Using the theory of regenerative processes we have with probability 1,

$$
\begin{aligned}
&\lim_{t\to\infty}\frac{Q(t)}{t}=\frac{E[Q]}{E[T]},\ \lim_{t\to\infty}\frac{B(t)}{V(t)}=\frac{E[B]}{E[V]},\ \lim_{t\to\infty}\frac{B_1(t)}{V(t)}=\frac{E[B_1]}{E[V]},\\
&\lim_{t\to\infty}\frac{B_2(t)}{V(t)}=\frac{E[B_2]}{E[V]},\ \lim_{t\to\infty}\frac{S(t)}{N(t)}=\frac{E[S]}{E[N]},\ \lim_{t\to\infty}\frac{S_1(t)}{N(t)}=\frac{E[S_1]}{E[N]},\\
&\lim_{t\to\infty}\frac{S_2(t)}{N(t)}=\frac{E[S_2]}{E[N]},\ \lim_{t\to\infty}\frac{J(t)}{t}=\frac{E[J]}{E[T]},\ \lim_{t\to\infty}\frac{C(t)}{t}=\frac{E[C]}{E[T]}.
\end{aligned}
\tag{3.1.4}
$$

Using the familiar relations $E[N] = \lambda E[T]$, $E[V] = \lambda E[D].E[T]$ and $E[J] = E[B_1]/\pi_2$, it remains to find tractable expressions for $E[T]$, $E[Q]$, $E[B_1]$, $E[B_2]$, $E[S_1]$, $E[S_2]$ and $E[C]$.

Towards this end we define as in the chapters 1 and 2 a number of basic functions. First we define the basic functions associated with production rate π_1. Assume that at epoch 0 the inventory level equals $x+m$, $x \geq 0$, and production rate π_1 is used. Then we define

$t_1(x)$:= the expected time until the inventory level decreases below m.

$$p(x,u) := \begin{cases} \text{the probability that the undershoot of level m is} \\ \text{greater than u when the inventory level decreases} \\ \text{below m for the first time, } 0 \leqq u \leqq m+L_0. \\ \\ \text{the probability that an amount of demand greater than} \\ u-m-L_0 \text{ is lost when the inventory level decreases below m} \\ \text{for the first time, } u > m+L_0 \end{cases}$$

Through the definition of $p(x,u)$ we again have that

$$(3.1.5) \qquad p(x,u) = p_B(x,u), \quad x \geq 0, \; u \geq 0,$$

where $p_B(x,u)$ denotes the distribution of the undershoot of level m in the backlog model. Also, using the principle of conservation of flow, we obtain

$$(3.1.6) \qquad t_1(x) = \frac{x + \int_0^\infty p(x,u)\,du}{\lambda E[D]-\pi_1}$$

Thus we find exact expressions for $t_1(0)$ and $p(0,u)$ and approximations for $t_1(x)$ and $p(x,u)$ for x sufficiently large.

Now we assume that at epoch 0 the inventory equals x, $-L_0 \leqq x \leqq M$, and production rate π_2 is used. Then we define the random variable

$T_2(x;M)$:= the time until the inventory reaches the level M,

and the basic functions

$t_2(x;M) := E[T_2(x;M)]$.

$q(x;M)$:= the probability that a stockout occurs in $(0,T_2(x;M)]$.

$n(x;M)$:= the expected number of stockout occurrences in $(0,T_2(x;M)]$.

$b_1(x;M)$:= the expected amount of excess demand in $(0,T_2(x;M)]$. that is backlogged.

$b_2(x;M)$:= the expected amount of excess demand in $(0,T_2(x;M)]$ that is lost.

$s_1(x;M)$:= the expected number of customers arriving in $(0,T_2(x;M)]$ whose demands cannot be met directly from stock on hand, but are ultimately satisfied completely.

$s_2(x;M)$:= the expected number of customers arriving in $(0,T_2(x;M)]$ whose demands are not ultimately satisfied.

$c(x;M)$:= the cumulative backlog at time $T_2(x;M)$.

Note that now we explicitly express that these basic functions depend on the value of M.

Then we find the following relations

$$E[T] = t_1(M-m) + \int_0^{m+L_0} t_2(m-u;M)\,d_u(1-p(M-m,u)) + t_2(-L_0;M)\,p(M-m,m+L_0). \tag{3.1.7}$$

$$E[Q] = \int_0^{m} n(m-u;M)\,d_u(1-p(M-m,u)) + (1+n(0;M))\,p(M-m,m). \tag{3.1.8}$$

$$E[B_1] = \int_0^{m+L_0} b_1(m-u;M)\,d_u(1-p(M-m,u)) + \int_m^{m+L_0} (u-m)\,d_u(1-p(M-m,u)) + (L_0+b_1(-L_0;M))\,p(M-m,m+L_0). \tag{3.1.9}$$

$$E[B_2] = \int_0^{m+L_0} b_2(m-u;M)\,d_u(1-p(M-m,u)) + \int_{m+L_0}^{\infty} (u-m-L_0)\,d_u(1-p(M-m,u)) + b_2(-L_0;M)\,p(M-m,m+L_0). \tag{3.1.10}$$

$$E[S_1] = \int_0^{m+L_0} s_1(m-u;M)\,d_u(1-p(M-m,u)) + p(M-m,m) - p(M-m,m+L_0) + s_1(-L_0;M)\,p(M-m,m+L_0). \tag{3.1.11}$$

$$E[S_2] = \int_0^{m+L_0} s_2(m-u;M)\,d_u(1-p(M-m,u)) + (1+s_2(-L_0;M))\,p(M-m,m+L_0). \tag{3.1.12}$$

$$E[C] = \int_0^{m+L_0} c(m-u;M)\,d_u(1-p(M-m,u)) + c(-L_0;M)\,p(M-m,m+L_0). \tag{3.1.13}$$

Thus the computation of E[T], E[Q], $E[B_1]$, $E[B_2]$, $E[S_1]$, $E[S_2]$ and E[C] can be reduced to the derivation of expressions for the above defined basic functions. We first note that we only need expressions for $t_1(x)$ and $p(x,u)$ for x=M-m. We distinguish between the cases M=m and M>m. In view of (3.1.5) and (3.1.6) we can state the following results.

Case M=m. In this case we can use the exact results given in section 1.3. Expressions for $t_1(0)$ and $p(0,u)$ are given by the equations (1.3.33) and (1.3.34) respectively.

Case M>m. In this case we use the approximations 1.3.1 and 1.3.2 for $t_1(M-m)$ and $p(M-m,u)$ respectively. To ensure the accuracy of these approximate expressions we assume that M-m satisfies condition 1.3.1.

In the next section we express $t_2(x;M)$, $q(x;M)$, $n(x;M)$, $b_1(x;M)$, $b_2(x;M)$, $s_1(x;M)$, $s_2(x;M)$ and $c(x;M)$ in terms of the basic functions for the backlog and lost-sales models. For the latter basic functions we derived expressions in chapter 1 and 2. Thus we obtain expressions for the basic functions associated with the present model with mixed backlogging and lost-sales.

3.2. The key relations.

In this section we give a number of exact relations between the basic functions of the present model and the basic functions of the backlog and lost-sales models. Throughout this section we assume that at epoch 0 production rate π_2 is used and, unless stated otherwise, the inventory $X(0)=x$ with $-L_0 \leq x \leq M$.

We define for the backlog model

$b_B(x;M)$:= the expected amount of demand backlogged until the inventory reaches the level M.

$q_B(x;M)$:= the probability that a stockout occurs before the inventory reaches the level M.

$c_B(x;M)$:= the cumulative backlog at the epoch at which the inventory reaches the level M for the first time.

We define for the lost-sales model, for all $0 \leq x \leq M$

$t_{2,L}(x;M)$:= the expected time until the inventory reaches the level M.

$b_L(x;M)$:= the expected amount of demand lost until the inventory reaches the level M.

$n_L(x;M)$:= the expected number of stockout occurrences until the inventory reaches the level M.

We note that the basic function $n_L(x;M)$ has not been defined in chapter 2. However, using the equation (2.3.1), which states that the probability of a stockout-occurrence before the inventory reaches the level M is the same for the backlog and lost-sales models, we can express $n_L(x;M)$ in terms of $q_B(x;M)$. We condition on the event of a stockout occurrence before the inventory reaches the level M. Then we obtain

$$(3.2.1) \qquad n_L(x;M) = q_B(x;M)(1+n_L(0;M)).$$

Setting x equal to 0 we can solve for $n_L(0;M)$,

$$n_L(0;M) = q_B(0;M)(1-q_B(0;M))^{-1}.$$

Substituting this result into (3.2.1), we find

$$(3.2.2) \qquad n_L(x;M) = q_B(x;M)(1-q_B(0;M))^{-1}.$$

Hence we obtain an approximation for $n_L(x;M)$ from equation (3.2.2) and approximation 1.4.1.

Now we can give the following set of key relations

$$(3.2.3) \qquad q(x;M) = q_B(x;M), \qquad 0 \leq x \leq M.$$

$$(3.2.4)\qquad n(x;M) = \begin{cases} n_L(x;M), & 0 \leq x \leq M \\ n_L(0;M), & -L_0 \leq x < 0 \end{cases}$$

$$(3.2.5)\qquad t_2(x;M) = t_{2,L}(x+L_0;M+L_0), \qquad -L_0 \leq x \leq M.$$

$$(3.2.6)\qquad b_2(x;M) = b_L(x+L_0;M+L_0), \qquad -L_0 \leq x \leq M.$$

$$(3.2.7)\qquad s_2(x;M) = n_L(x+L_0;M+L_0), \qquad -L_0 \leq x \leq M.$$

$$(3.2.8)\qquad b_1(x;M) = b_B(x;M)-b_B(x+L_0;M+L_0), \qquad -L_0 \leq x \leq M.$$

$$(3.2.9)\qquad s_1(x;M) = n(x;M)+\frac{\lambda b_1(x;M)}{\pi_2} - s_2(x;M), \qquad -L_0 \leq x \leq M.$$

$$(3.2.10)\qquad c(x;M) = c_B(x;M)-c_B(x+L_0;M+L_0)-\frac{L_0}{\pi_2} b_B(x+L_0;M+L_0), \qquad -L_0 \leq x \leq M.$$

Using these key relations in the right order we have indeed expressed the basic functions for the current model in terms of basic functions, for which tractable expressions are available. We shall only give an outline of the intuitive ideas behind the proof of the above relations.

Let us first consider the relations (3.2.3) and (3.2.4). Relation (3.2.3) follows from the fact that given $X(0)\geq 0$ the sample paths in the present model and the backlog model are identical *as long as the inventory level is non-negative*. By the same arguments as used to derive (3.2.2) relation (3.2.4) then follows from (3.2.3). Here we also use that given $X(0)<0$ the inventory will eventually reach the level 0 when production rate π_2 is used.

To justify relations (3.2.5) to (3.2.7) we shall use a sample path argument. Let us define the random variables

τ_1 := the arrival epoch of the first customer.

τ_n := the time that elapses between the arrival of the (n-1)-th and n-th customer, $n \geq 2$.

D_n := the demand of the n-th arriving customer, $n \geq 1$.

The sequences $\{\tau_n\}$ and $\{D_n\}$ are defined on some common probability space. Let ω be an element of the underlying sample space and $\{\tau_n(\omega)\}$ and $\{D_n(\omega)\}$ the realizations of $\{\tau_n\}$ and $\{D_n\}$ for this ω. From the sequences $\{\tau_n(\omega)\}$ and $\{D_n(\omega)\}$ we construct the sample path for two models:

model I. $X(0)=x$, production rate π_2 is used, switch level M and a maximum backlog of L_0.

model II. $X(0)=x+L_0$, production rate π_2 is used, switch level $M+L_0$ and no backlog allowed.

The model I corresponds to the mixed backlogging and lost-sales model, while model II is a lost-sales model as studied in chapter 2. Let us define

$X^I(t)(\omega)$ ($X^{II}(t)(\omega)$) = the inventory level at time t in model I (II) for the given sample point ω.

Then it is immediately clear that

$$X^I(t)(\omega) = X^{II}(t)(\omega)-L_0, \quad 0 \leq t \leq T_2^I(x)(\omega).$$

Here $T_2^I(x)(\omega)$ is the realization of $T_2^I(x)$ for the sample point ω and $T_2^I(x)$ is defined analogously to $T_2(x;M)$ in section 3.1. Hence for all sample points ω in the sample space the time until the inventory reaches the level M in model I equals the time until the inventory reaches the level $M+L_0$ in model II. There are similar equalities for the amount of demand lost in the models I and II and for the number of customers whose demands are partially lost in the models I and II, yielding the equations (3.2.5)-(3.2.7).

Next we motivate equation (3.2.8). Let us consider the backlog model of chapter 1. Given that at epoch 0 the inventory level equals $-L_0 \leq x \leq M$ and production rate π_2 is used, we define the random variables

$T_{2,B}(x)$:= the time until the inventory level reaches the value M,

$B_B(x)$:= the amount of demand backlogged during $(0, T_2(x)]$ (excluding shortages existing at epoch 0),

$$B_B'(x) := \sum_{n=1}^{N_B(T_{2,B}(x))} \; \prod_{\{X_B(\sum_{j=1}^{n}\tau_j) < -L_0\}} [\min(-L_0, X_B((\sum_{j=1}^{n}\tau_j)^-)) - X_B(\sum_{j=1}^{n}\tau_j)],$$

$$B_B''(x) := B_B(x) - B_B'(x).$$

Here $N_B(t)$ and $X_B(t)$ for the backlog model are defined analogously to $N(t)$ and $X(t)$ for the present model. Also, $X_B(t)$ is right-continuous.

To elucidate the definition of $B_B'(x)$ and $B_B''(x)$ we consider figure 3.2.1.

FIGURE 3.2.1.

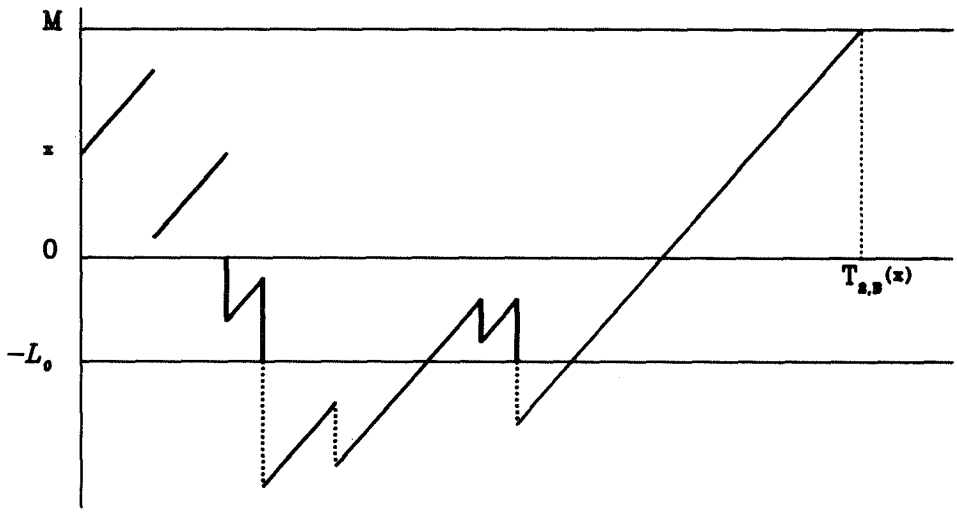


Figure 3.2.1 shows a typical sample path for the backlog model. Now $B'_B(x)$ equals the total length of the dotted lines, while $B''_B(x)$ equals the total length of the solid lines.

By definition, we have that

$$(3.2.11) \qquad b_B(x;M) = E[B_B(x)], \quad -L_0 \leq x \leq M.$$

By the same sample path argument as we used to proof equations (3.2.5)-(3.2.7) we now obtain

$$(3.2.12) \qquad E[B'_B(x)] = b_B(x+L_0;M+L_0), \quad -L_0 \leq x \leq M.$$

To determine $E[B''_B(x)]$ we note that when in the backlog model the inventory drops below $-L_0$ then with probability 1 the inventory will reach the level $-L_0$ from below. Now it follows from the Markov property of the exponential interarrival times that each time the level $-L_0$ is reached an exponentially distributed time elapses until the next customer arrives. This special property is the key to the following statement:

> By deleting the time intervals during which the inventory level is below $-L_0$, we construct a new process, whose probabilistic behaviour is identical to the inventory process in the model with a maximum backlog of L_0.

It will now be clear that

$$(3.2.13) \qquad E[B''_B(x)] = b_1(x;M).$$

This result can be proved rigorously by sample path arguments. Combining the definitions of $B_B(x)$, $B'_B(x)$ and $B''_B(x)$ with equations (3.2.11)-(3.2.13) we obtain equation (3.2.8).

The equation (3.2.9) can be derived as follows. We want to find an expression for $s_1(x;M)$. We observe that

(3.2.14) $s_1(x;M)$ = (the expected number of customers arriving in $(0,T_2(x;M)]$ whose demands cannot be met directly from stock on hand, but are ultimately satisfied completely)

= (the expected number of customers arriving in $(0,T_2(x;M)]$ whose demands cannot be met directly from stock on hand)
- (the expected number of customers arriving in $(0,T_2(x;M)]$ whose demands are partially lost).

The second term on the right-hand side of (3.2.14) equals $s_2(x;M)$. The first term on the right-hand side of (3.2.14) can be written as the sum of two expressions,

(3.2.15) (the expected number of customers arriving in $(0,T_2(x;M)]$ whose demands cannot be met directly from stock on hand)

= (the expected number of customers arriving in $(0,T_2(x;M)]$ that cause a stockout)
· (the expected number of customers arriving in $(0,T_2(x;M)]$ while the inventory is negative).

The first term on the right-hand side of (3.2.15) equals $n(x;M)$. An expression for the second term on the right-hand side of (3.2.15) can be found as follows. Since every backlog is produced at rate π_2 it follows that

(the expected time during $(0,T_2(x;M)]$ that the inventory is negative)
$= b_1(x;M)/\pi_2$.

Using "Poisson arrivals see time averages" it can be derived that

(3.2.16) (the expected number of customers arriving in $(0,T_2(x;M)]$, while the inventory is negative)
$= \frac{\lambda}{\pi_2} b_1(x;M)$.

Thus equations (3.2.14)-(3.2.16) together yield (3.2.9).

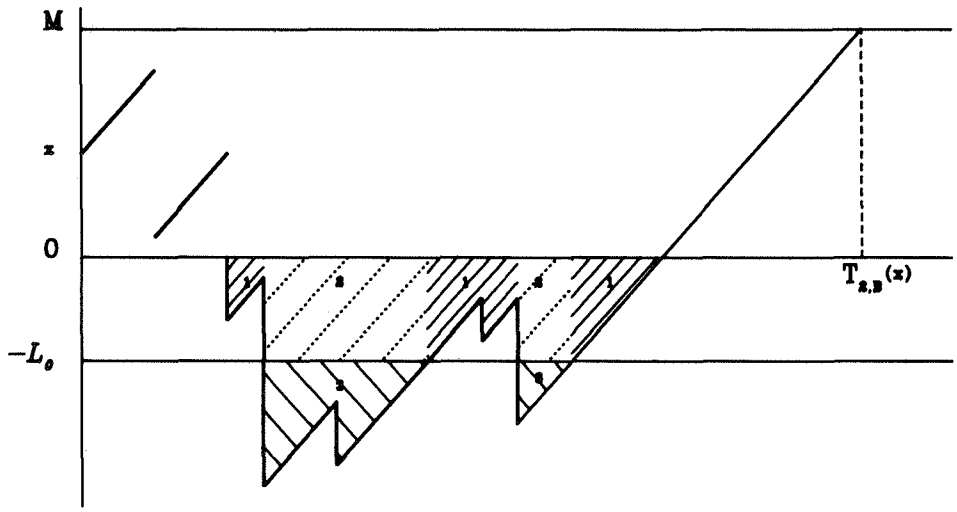
Finally we want to proof relation (3.2.10). As noted in chapter 1 it is helpful to give the cumulative backlog a cost interpretation by imagining that a penalty cost at rate z is incurred when a shortage of size z exists. Then

$$c(x;M) = \text{the expected penalty cost incurred in } (0,T_2(x;M],$$
$$-L_0 \leqq x \leqq M.$$

We shall express c(x;M) in terms of basic functions associated with the backlog model.

It is useful to consider figure 3.2.2 in which a typical sample path for the backlog model is drawn.

FIGURE 3.2.2.



We have hatched 3 areas in figure 3.2.2. By introducing an appropriate cost structure we can easily compute the expected values of these areas.

We first consider the shaded area 1 . Let us assume that a cost at rate z is incurred if a shortage of size z with $0 \leqq z \leqq L_0$ exists, otherwise no cost is incurred. For this particular cost structure, let

$$c_{1,B}(x;M) := \text{the expected cost incurred in } (0,T_{2,B}(x)],$$
$$-L_0 \leqq x \leqq M.$$

Then $c_{1,B}(x;M)$ is the expected value of area 1 . By the same arguments as we used to proof (3.2.13) we now find

$$(3.2.17) \qquad c_{1,B}(x;M) = c(x;M), \quad -L_0 \leq x \leq M.$$

Next we introduce another cost structure. We assume that a cost at rate L_0 is incurred if a shortage of size $z \geq -L_0$ exists, otherwise no cost is incurred. Under this cost structure, let

$$c_{2,B}(x;M) := \text{the expected shortage cost incurred in } (0,T_{2,B}(x)], \quad -L_0 \leq x \leq M.$$

Then $c_{2,B}(x;M)$ equals the expected value of area 2 . It easily follows that for all $-L_0 \leq x \leq M$

$$(3.2.18) \qquad c_{2,B}(x;M) = L_0 \times (\text{the expected time during } (0,T_{2,B}(x)] \text{ that the inventory level is below } -L_0).$$

By the sample path argument that was used to derive (3.2.5)-(3.2.7) and using the fact that every backlog is produced at rate π_2, we obtain for all $-L_0 \leq x \leq M$

$$(3.2.19) \qquad (\text{the expected time during } (0,T_{2,B}(x)], \text{ that the inventory is below } -L_0) = b_B(x+L_0;M+L_0)/\pi_2.$$

A combination of (3.2.18) and (3.2.19) yields

$$(3.2.20) \qquad c_{2,B}(x;M) = \frac{L_0}{\pi_2} . b_B(x+L_0;M+L_0), \quad -L_0 \leq x \leq M.$$

Now we introduce the following cost structure. A shortage cost at rate $z-L_0$ is incurred if a shortage of z exists with $z \geq L_0$, otherwise no cost is incurred. Under this cost structure, let

$$c_{3,B}(x;M) := \text{expected cost incurred in } (0,T_{2,B}(x)], \quad -L_0 \leq x \leq M.$$

As before we have chosen the cost structure such that $c_{3,B}(x;M)$ equals the expected value of area 3 . The sample path argument used to derive (3.2.19) now implies that

$$(3.2.21) \qquad c_{3,B}(x;M) = c_B(x+L_0;M+L_0), \quad -L_0 \leq x \leq M.$$

We have determined the expected values of area 1 , 2 and 3 . It is immediately clear that the sum of these areas corresponds to the cost incurred under the following cost structure: A cost is incurred at rate z if a backlog of size z exists. This is equivalent to assuming that the three particular costs considered above are incurred simultaneously. It follows from the cost interpretation of the cumulative backlog that

$$c_B(x;M) = \text{the expected cost incurred in } (0,T_{2,B}(x)] \text{ if the three particular costs are incurred simultaneously}$$

or equivalently

$$(3.2.22) \qquad c_B(x;M) = c_{1,B}(x;M)+c_{2,B}(x;M)+c_{3,B}(x;M), \quad -L_0 \leq x \leq M.$$

Thus it follows from (3.2.17), (3.2.20), (3.2.21) and (3.2.22) that

$$c_B(x;M) = c(x;M)+\frac{L_0}{\pi_2} b_1(x+L_0,M+L_0)+c_B(x+L_0,M+L_0), \quad -L_0 \leq x \leq M.$$

Rearranging terms leads to equation (3.2.10).

We have now expressed the basic functions associated with the production rates π_1 and π_2 in terms of basic functions for the backlog and lost-sales model. It is now a matter of combining the relations (3.1.5)-(3.1.13) and (3.2.3)-(3.2.10) with the approximations 1.3.1, 1.3.2, 1.4.1-1.4.3 and 2.3.1-2.3.3 to obtain expressions for the service measures defined in section 3.1.

Because of the fact that the relations (3.1.5), (3.2.2)-(3.2.10) are exact, the accuracy of the approximations for the model considered in this chapter follows from the established accuracy of the approximations in the chapters 1 and 2. We refer to sections 1.5 and 2.4 for further comments.

Remark 3.2.1. For the case of exponential demand the explicit and implicit approximations given in this chapter are exact. This follows from the fact that the approximations obtained in chapters 1 and 2 are exact for exponential demand and from the exact relations (3.1.5), (3.2.2)-(3.2.10).

Remark 3.2.2. When production rate $\pi_1=0$, and the inventory is controlled by an (m,M)-rule with M=m=0 then the process $\{-X(t),\ t\geq 0\}$ corresponds to the workload process in an M/G/1-dam model with processing rate π_2. From the results obtained in this chapter we can derive approximations for the fraction of work that is lost, the fraction of customers whose workloads are not processed completely and the average workload.

4. THE SINGLE PRODUCT PRODUCTION-INVENTORY MODEL IN WHICH EXCESS DEMAND IS EITHER BACKLOGGED OR COMPLETELY LOST.

In chapter 3 we assumed that the backlog at any point in time cannot be larger than a given amount L_0 so that customer demands by which the backlog would exceed this level are partially lost. In this chapter we allow backlogs larger than L_0 but we now make the stipulation that any demand occurring while a backlog larger than L_0 exists is completely lost. A demand occurring, when no backlog exists or the backlog is not larger than L_0, is ultimately satisfied completely. Note that a demand occurring in the latter situation may cause the backlog to exceed L_0.

As in chapter 3 we will derive expressions for a number of service measures, where we use results that have been derived in earlier chapters. In particular we will find exact relations between basic functions for the model described here and the model discussed in chapter 3.

4.1. Model and preliminaries.

At a production facility manufacturing a single commodity customers arrive according to a Poisson process with rate λ. The demands of these customers for the commodity are independent random variables having a common distribution function F with $F(0)=0$ and $1-F(x)=0(e^{-\kappa x})$ as $x\to\infty$ for some $\kappa>0$. The demand sizes are independent of the arrival process.

The commodity is continually produced by using one out of two rates π_1 and π_2, such that

$$(4.1.1) \qquad \pi_1 < \lambda E[D] < \pi_2.$$

The generic random variable D denotes the demand of a single customer. Control on the inventory is exercized according to an (m,M)-rule with $0\leq m\leq M$. The system has an infinite storage capacity.

Any customer finding upon arrival a backlog larger than a given amount $L_0\geq 0$ leaves immediately and the demand of such a reneging customer is completely lost. Otherwise the demand of an arriving customer is ultimately satisfied completely immediately and/or by later production. Then a customer arrving, when no backlog exists or the backlog is not larger than L_0, may cause the backlog to exceed L_0.

As before we focus on service measures like the long-run fraction of demand lost. We will express these service measures in terms of a number of basic functions. We will derive exact relations between these basic functions and the basic functions associated with the model discussed in chapter 3.

The random variables and basic functions we use below are defined as in section 3.1. We derive relations between the following two models,

model I: the current production-inventory model where customers immediately leave the system, if at the time of their arrival the backlog exceeds L_0.

model II: the previous production-inventory model with mixed backlogging and lost-sales. The maximal backlog is L_0.

To distinguish between the random variables and basic functions of these two models we use the superscripts I and II.

First of all we will make the following important observation. When the inventory level is not below $-L_0$ the inventory processes behave probabilistically identically in the models I and II. As usual, this observation together with the lack of memory of the Poisson arrival process will be essential in our analysis. The results to be presented below will be derived in a somewhat informal way in order not to obscure the ideas behind the proofs. However, the proofs can be made rigorously by using rather well-known sample path arguments.

The inventory process in model I is regenerative. Therefore we can express the service measures describing the long-run behaviour of the system in terms of expected values of random variables corresponding to the behaviour of the system during one cycle. We again consider the following service measures.

(i) <u>α-service measure.</u>

the long-run average number of stockouts per unit time,

$$\lim_{t\to\infty} \frac{Q^I(t)}{t} = \frac{E[Q^I]}{E[T^I]} \text{ with probability 1.}$$

(ii) β-service measure.

the long-run fraction of demand that cannot be met directly from stock on hand,

$$\lim_{t\to\infty} \frac{B^I(t)}{V^I(t)} = \frac{E[B^I]}{E[V^I]} \text{ with probability 1}$$

(iii) β_1-service measure.

the long-run fraction of demand that is backlogged,

$$\lim_{t\to\infty} \frac{B_1^I(t)}{V^I(t)} = \frac{E[B_1^I]}{E[V^I]} \text{ with probability 1.}$$

(iv) β_2-service measure.

the long-run fraction of demand that is lost,

$$\lim_{t\to\infty} \frac{B_2^I(t)}{V^I(t)} = \frac{E[B_2^I]}{E[V^I]} \text{ with probability 1.}$$

(v) γ-service measure.

the long-run fraction of customers whose demands cannot be met directly from stock on hand,

$$\lim_{t\to\infty} \frac{S^I(t)}{N^I(t)} = \frac{E[S^I]}{E[N^I]} \text{ with probability 1.}$$

(vi) γ_1-service measure.

the long-run fraction of customers whose demands cannot be met directly from stock on hand, but are ultimately satisfied completely,

$$\lim_{t\to\infty} \frac{S_1^I(t)}{N^I(t)} = \frac{E[S_1^I]}{E[N^I]} \text{ with probability 1.}$$

(vii) γ_2-service measure.

the long-run fraction of customers whose demands are not ultimately satisfied and are completely lost,

$$\lim_{t\to\infty} \frac{S_2^I(t)}{N^I(t)} = \frac{E[S_2^I]}{E[N^I]} \text{ with probability 1.}$$

(viii) δ-service measure.

the long-run average backlog at an arbitrary point in time,

$$\lim_{t\to\infty} \frac{C^I(t)}{t} = \frac{E[C^I]}{E[T^I]} \text{ with probability 1.}$$

Using the relations $E[N^I]=\lambda E[T^I]$ and $E[V^I]=\lambda E[D]E[T^I]$ it follows that we must find expressions for $E[T^I]$, $E[Q^I]$, $E[B_1^I]$, $E[B_2^I]$, $E[S_1^I]$, $E[S_2^I]$ and $E[C^I]$.

We express these quantities in terms of basic functions. These basic functions are defined as in section 3.1, with exception of

$p^I(x,u)$:= the probability that the undershoot of level m is greater than u when the inventory level decreases below m for the first time, given $X(0)=x+m$ and at epoch 0 production rate π_1 is used.

Also, we note that for the present model the backlog at an arbitrary point in time is unbounded. This implies that the basic functions associated with production rate π_2 have to be defined for any initial inventory $X(0)=x$ with $x\leq M$. We obtain the following relations.

$$E[T^I] = t_1^I(M-m) + \int_0^\infty t_2^I(m-u;M)\,d_u(1-p^I(M-m,u)). \tag{4.1.2}$$

$$E[Q^I] = \int_0^m n^I(m-u;M)\,d_u(1-p^I(M-m,u)) + (1+n^I(0;M))p^I(M-m,m). \tag{4.1.3}$$

$$E[B_1^I] = \int_0^\infty b_1^I(m-u;M)\,d_u(1-p^I(M-m,u)) + \int_m^\infty (u-m)\,d_u(1-p^I(M-m,u)). \tag{4.1.4}$$

$$E[B_2^I] = \int_0^\infty b_2^I(m-u;M)\,d_u(1-p^I(M-m,u)). \tag{4.1.5}$$

$$E[S_1^I] = \int_0^\infty s_1^I(m-u;M)\,d_u(1-p^I(M-m,u)) + p^I(M-m,m). \tag{4.1.6}$$

$$E[S_2^I] = \int_0^\infty s_2^I(m-u;M)\,d_u(1-p^I(M-m,u)). \tag{4.1.7}$$

$$E[C^I] = \int_0^\infty c^I(m-u;M)\,d_u(1-p^I(M-m,u)). \tag{4.1.8}$$

In the next section we will derive a number of exact relations between the basic functions associated with model I and those associated with model II as well as the backlog model. These relations then yield tractable expressions for the above defined service measures.

4.2. The key relations.

We can give the following relations between the basic functions associated with model I and those associated with model II and the backlog model. The basic functions for the backlog model, identifyable by the subscript "B", are defined as in section 1.2.

$$t_1^I(x) = t_{1,B}(x), \qquad x \geq 0 \tag{4.2.1}$$

$$p^I(x,u) = p_B(x,u), \qquad x \geq 0,\ u \geq 0 \tag{4.2.2}$$

$$n^I(x;M) = n^{II}(x;M), \qquad 0 \leq x \leq M \tag{4.2.3}$$

$$b_1^I(x;M) = b_1^{II}(x;M) + b_2^{II}(x;M), \qquad -L_0 \leq x \leq M \tag{4.2.4}$$

$$s_1^I(x;M) = s_1^{II}(x;M) + s_2^{II}(x;M), \qquad -L_0 \leq x \leq M \tag{4.2.5}$$

$$b_1^I(x;M) = b_1^I(-L_0;M), \qquad x < -L_0 \tag{4.2.6}$$

$$s_1^I(x;M) = s_1^I(-L_0,M), \qquad x < -L_0 \tag{4.2.7}$$

$$t_2^I(x;M) = \begin{cases} t_2^{II}(x;M) + b_2^{II}(x;M)/\pi_2 & -L_0 \leq x \leq M \\ (-L_0-x+b_2^{II}(-L_0;M))/\pi_2 + t_2^{II}(-L_0;M), & x < -L_0 \end{cases} \tag{4.2.8}$$

$$b_2^I(x;M) = \begin{cases} \frac{\lambda E[D]}{\pi_2}\ b_2^{II}(x;M) & -L_0 \leq x \leq M \\ \frac{\lambda E[D]}{\pi_2}\ (-L_0-x+b_2^{II}(-L_0;M)), & x < -L_0 \end{cases} \tag{4.2.9}$$

$$s_2^I(x;M) = \begin{cases} \frac{\lambda}{\pi_2}\ b_2^{II}(x;M) & -L_0 \leq x \leq M \\ \frac{\lambda}{\pi_2}\ (-L_0-x+b_2^{II}(-L_0;M)) & x < -L_0 \end{cases} \tag{4.2.10}$$

$$(4.2.11)\qquad c^{I}(x;M) = c^{II}(x;M) + \frac{L_0}{\pi_2}\, b_2^{II}(x;M) + (1-\frac{\lambda E[D]}{\pi_2})c_B(x+L_0;M+L_0)$$
$$- \frac{\lambda E[D^2]}{2\pi_2^2}\, b_B(x+L_0;M+L_0), \qquad -L_0 \leq x \leq M$$

$$(4.2.12)\qquad c^{I}(x;M) \quad \frac{(x^2-L_0^2)}{2\pi_2} + c^{I}(-L_0;M), \qquad x < -L_0$$

Let us consider equations (4.2.1) and (4.2.2). If $X^I(0)=X_B(0)=x+m$ with $x\geqq 0$ and at epoch 0 a production rate π_1 is used then sample paths for model I and the backlog model are identical until the inventory becomes negative for the first time. In particular the inventory position immediately after an undershoot of the level m and the epoch at which this is occurring are the same in both models. Thus equations (4.2.1) and (4.2.2) hold.

From now on we assume that $X^I(0)=x\leqq M$ and at epoch 0 a production rate π_2 is used.

Equation (4.2.3) immediately follows from the observation made in section 4.1 that model I and model II are probabilistically identical if these processes are observed only at time intervals at which the inventory is above level $-L_0$.

Next we consider equations (4.2.4) and (4.2.5). Let $X^I(0)=X^{II}(0)=x$ with $-L_0\leqq x\leqq M$. Then it is immediately clear that

$$(4.2.13)\qquad b_1^I(x;M) = \text{(the expected amount of demand backlogged, during the time that the inventory level is above } -L_0 \text{ until the inventory reaches the level M)}$$
$$+ \text{ (the expected amount of demand backlogged during the time that the inventory level is below } -L_0 \text{ until the inventory reaches the level M).}$$

Since the models I and II behave identically when the inventory is not below $-L_0$ it follows that the first expression between brackets on the right-hand side of (4.2.13) equals $b_1^{II}(x;M)$. Some reflections show that this observation also implies that the second term on the right-hand side of (4.2.13) equals $b_2^{II}(x;M)$. Putting it more precisely, the amounts of lost demands of arriving customers in model II correspond to the undershoots of the level $-L_0$ in model I. Thus we obtain equation (4.2.4). The same arguments can be applied to obtain equation (4.2.5).

Let us assume that $X^I(0)=x<-L_0$ and at epoch 0 the production rate π_2 is used. In the definition of $b_1^I(x)$ we exclude the existing backlog of $-x$ at epoch 0. Since customers finding upon arrival the inventory below $-L_0$ immediately leave the system, it follows that the demands of the customers arriving in $(0,(-L_0-x)/\pi_2]$ are lost. Then it follows from the lack of memory of the Poisson arrival process that the expected amount of demand backlogged during $((-L_0-x)/\pi_2, T_2^I(x;M)]$ equals $b_1^I(-L_0;M)$. This implies equation (4.2.6) and similarly we obtain equation (4.2.7).

Now we derive equations (4.2.8)-(4.2.10). First we define

$$\bar{t}(x;M) := \text{the expected time that the inventory is below } -L_0 \text{ until the inventory reaches the level M.}$$

Then it follows from the important observation that model I and model II are probabilistically identical if these processes are observed only at times when the inventory is above $-L_0$, that

$$(4.2.14) \qquad t_2^I(x;M) = t_2^{II}(x;M) + \bar{t}(x;M), \quad -L_0 \leq x \leq M.$$

As said before, the amount of demand backlogged during the time that the inventory is below the level $-L_0$ equals $b_2^{II}(x;M)$. Since any backlog is produced at rate π_2 we obtain

$$(4.2.15) \qquad \bar{t}(x;M) = b_2^{II}(x;M)/\pi_2, \quad -L_0 \leq x \leq M$$

$$(4.2.16) \qquad \bar{t}(x;M) = (-L_0-x)/\pi_2 + \bar{t}(-L_0;M), \quad x < -L_0.$$

In equation (4.2.16) we also used the lack of memory of the Poisson arrival process and the fact that during $(0,(-L_0-x)/\pi_2]$, with $x\leq -L_0$, the demands of arriving customers are lost. Equations (4.2.14), (4.2.15) and (4.2.16) together imply equation (4.2.8).

Since customers arriving during the time that the inventory is below $-L_0$ immediately leave the system and using that the demand process is a compound Poisson process, it follows that

$$(4.2.17) \qquad b_2^I(x;M) = \lambda E[D]\bar{t}(x;M)$$

(4.2.18) $\quad s_2^I(x;M) = \lambda t^-(x;M).$

Combining equations (4.2.15)-(4.2.18) we obtain the equations (4.2.9) and (4.2.10).

We finally prove equations (4.2.11) and (4.2.12). We recall that $c^I(x;M)$, the expected cumulative backlog at epoch $T_2^I(x;M)$ has the following cost interpretation. Assume that a cost at rate z is incurred when a backlog of size z exists. Then we have

$$c^I(x;M) = \text{the expected cost incurred in } (0,T_2^I(x;M)] \text{ when the initial inventory is } x, \quad x \leq M.$$

FIGURE 4.2.1.

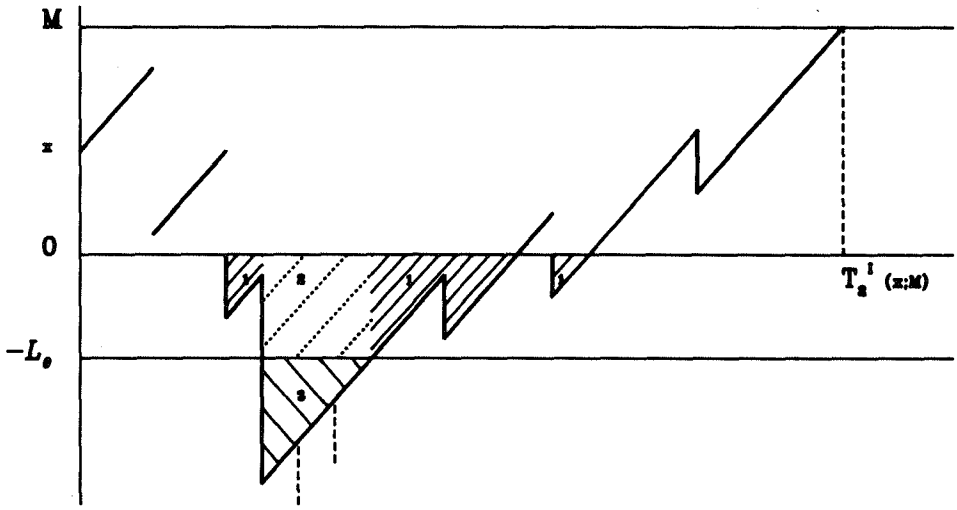


Figure 4.2.1 shows a typical sample path of the process $\{X^I(T),\ t\geq 0\}$ with $0\leq t\leq T_2^I(x;M)$ and $X^I(0)=x$ with $-L_0\leq x\leq M$. The dotted lines correspond to lost demands. The penalty cost incurred during $(0,T_2^I(x;M)]$ for this particular sample path equals the sum of areas 1 , 2 and 3 . We will calculate each of these areas by using an appropriately chosen cost structure. The discussion below assumes that production rate π_2 is used.

Let us first consider the shaded area 1 . Let us assume that a cost is incurred at rate z if a shortage of size z exists with $0\leq z\leq L_0$. Otherwise no cost is incurred. Under this cost structure, let

$$c_1^I(x;M) := \text{the expected cost incurred in } (0,T_2^I(x;M)],$$
$$-L_0 \leqq x \leqq M.$$

Then $c_1^I(x;M)$ equals the expected value of area 1. On the other hand, since the inventory processes in the models I and II behave probabilistically identically when the inventory level is not below $-L_0$, it follows that

$$(4.2.19) \qquad c_1^I(x;M) = c^{II}(x;M), \quad -L_0 \leqq x \leqq M.$$

Next we consider another cost structure. A cost is incurred at rate L_0 if a backlog exists of size $z \geqq L_0$, otherwise no cost is incurred. Under this cost structure, let

$$c_2^I(x;M) := \text{the expected penalty cost incurred in } (0,T_2^I(x;M)]$$
$$-L_0 \leqq x \leqq M.$$

It is immediately clear that $c_2^I(x;M)$ equals the expected value of the dotted area 2. Since a cost at rate L_0 is incurred during the time that the inventory is below $-L_0$, it follows from the definition of $t^-(x;M)$ that

$$c_2^I(x;M) = L_0\, t^-(x;M).$$

Using equation (4.2.15) we obtain

$$(4.2.20) \qquad c_2^I(x;M) = \frac{L_0}{\pi_2}\, b_2^{II}(x;M).$$

To obtain an expression for the expected value of area 3 we introduce the following cost structure. A cost is incurred at rate $z-L_0$ if a backlog exists of size $z > L_0$, otherwise no cost is incurred. Under this cost structure, let

$$c_3^I(x;M) := \text{the expected cost incurred in } (0,T_2^I(x;M)]$$
$$-L_0 \leqq x \leqq M.$$

Shifting each sample path of $\{X^I(t),\ t \geqq 0\}$ vertically upwards by an amount L_0 we find

$$(4.2.21) \qquad c_3^I(x;M) = c_0^I(x+L_0;M+L_0), \quad -L_0 \leq x \leq M$$

where

$$c_0^I(x;M) := \text{the expected cumulative backlog at time } T_2^I(x;M) \text{ when } L_0 = 0.$$

Hence it suffices to find an expression for $c_0^I(x;M)$ for $0 \leq x \leq M$.

Let us assume that $L_0=0$. Conditioning on the event of an arrival in $(0,\Delta x/\pi_2]$ it follows by standard arguments that

$$c_0^I(x;M) = (1-\frac{\lambda\Delta x}{\pi_2})c_0^I(x+\Delta x;M) + \frac{\lambda\Delta x}{\pi_2}\{\int_0^x c_0^I(x-y;M)dF(y)$$

$$+ c_0^I(0;M)(1-F(x)) + \int_x^\infty (y-x)^2/(2\pi_2)dF(y)\} + o(\Delta x),$$

$$x \geq 0.$$

Letting $\Delta x \to 0$ we obtain for almost all $0 \leq x \leq M$

$$(4.2.22) \qquad \frac{d}{dx} c_0^I(x;M) = -\frac{\lambda}{\pi_2} c_0^I(0;M)(1-F(x)) - \frac{\lambda}{2\pi_2^2}\int_x^\infty (y-x)^2 dF(y)$$

$$+ \frac{\lambda}{\pi_2} c_0^I(x;M) - \frac{\lambda}{\pi_2}\int_0^x c_0^I(x-y;M)dF(y).$$

Equation (4.2.22) holds for all M. Setting M equal to ∞ we can rewrite (4.2.22) into a defective renewal equation. Thus (4.2.22) has a unique solution $c_0^I(x;\infty)$ with $x \geq 0$ for the case of $M=\infty$. Using the lack of memory of the Poisson arrival process, it can be seen that

$$c_0^I(x;M) = c_0^I(x;\infty) - c_0^I(M;\infty).$$

Using $c_0^I(M;M)=0$ this result and the uniqueness of $c_0^I(x;\infty)$ imply that (4.2.22) has a unique solution $c_0^I(x;M)$ with $0 \leq x \leq M$. Below we construct this solution using integro-differential equations previously derived for the backlog model.

We recall the definitions of the basic functions $b_B(x;M)$ and $c_B(x;M)$ for the backlog model given in section 1.2. Given that at epoch 0 the inventory equals $0 \leq x \leq M$ and production rate π_2 is used we define

$b_B(x;M)$:= the expected amount of demand backlogged until the inventory reaches the level M.

$c_B(x;M)$:= the cumulative backlog at the epoch at which the inventory reaches the level M for the first time.

It follows from equations (1.4.20) and (1.4.38) that

$$(4.2.23)\quad \frac{d}{dx}\left((1-\frac{\lambda E[D]}{\pi_2})c_B(x;M)\right) = -\frac{\lambda}{\pi_2}(1-\frac{\lambda E[D]}{\pi_2})c_B(0;M)(1-F(x)) - \frac{\lambda}{2\pi_2^2}\int_x^\infty (y-x)^2 dF(y) - \frac{\lambda^2 E[D^2]}{2\pi^2(\pi_2-\lambda E[D])}\int_x^\infty (y-x)dF(y) - \frac{\lambda}{\pi_2}\int_0^x (1-\frac{\lambda E[D]}{\pi_2})c_B(x-y;M)dF(y), + \frac{\lambda}{\pi_2}(1-\frac{\lambda E[D]}{\pi_2})c_B(x;M), \quad 0 \leq x \leq M.$$

$$(4.2.24)\quad \frac{d}{dx}\left(\frac{\lambda E[D^2]}{2\pi_2^2} b_B(x;M)\right) = \frac{-\lambda^2 E[D^2]}{2\pi_2^2(\pi_2-\lambda E[D])}\int_x^\infty (y-x)dF(y) - \frac{\lambda}{\pi_2}\left(\frac{\lambda E[D^2]}{2\pi_2^2} b_B(0;M)\right)(1-F(x)) + \frac{\lambda}{\pi_2}\left(\frac{\lambda E[D^2]}{2\pi_2^2} b_B(x;M)\right) - \frac{\lambda}{\pi_2}\int_0^x \frac{\lambda E[D^2]}{2\pi_2^2} b_B(x-y;M)dF(y), \quad 0 \leq x \leq M.$$

Subtracting equation (4.2.24) from (4.2.23) we obtain

$$(4.2.25)\quad \frac{d}{dx}\left((1-\frac{\lambda E[D]}{\pi_2})c_B(x;M) - \frac{\lambda E[D^2]}{2\pi_2^2} b_B(x;M)\right) = -\frac{\lambda}{2\pi_2^2}\int_x^\infty (y-x)^2 dF(y) - \frac{\lambda}{\pi_2}\left[(1-\frac{\lambda E[D]}{\pi_2})c_B(0;M) - \frac{\lambda E[D^2]}{2\pi_2^2} b_B(0;M)\right](1-F(x)) + \frac{\lambda}{\pi_2}\left[(1-\frac{\lambda E[D]}{\pi_2})c_B(x;M) - \frac{\lambda E[D^2]}{2\pi_2^2} b_B(x;M)\right] - \frac{\lambda}{\pi_2}\int_0^x \left[(1-\frac{\lambda E[D]}{\pi_2})c_B(x-y;M) - \frac{\lambda E[D^2]}{2\pi_2^2} b_B(x-y;M)\right]dF(y), \quad 0 \leq x \leq M.$$

Comparing the equations (4.2.22) and (4.2.25) and using the uniqueness of the solution $c_0^I(x;M)$ to (4.2.22) we have that

$$(4.2.26) \qquad c_0^I(x;M) = (1-\frac{\lambda E[D]}{\pi_2})c_B(x;M) - \frac{\lambda E[D^2]}{2\pi_2^2}\, b_B(x;M), \quad 0 \leqq x \leqq M.$$

Then the equations (4.2.21) and (4.2.26) together imply

$$(4.2.27) \qquad c_3^I(x,M) = (1-\frac{\lambda E[D]}{\pi_2})c_B(x+L_0;M+L_0) - \frac{\lambda E[D^2]}{2\pi_2^2}\, b_B(x+L_0;M+L_0).$$

Using $c^I(x;M)=c_1^I(x;M)+c_2^I(x;M)+c_3^I(x;M)$ and combining (4.2.19), (4.2.20) and (4.2.27) we find equation (4.2.11).

Finally we prove equation (4.2.12). We assume that $X(0)=x<-L_0$. In the interval $(0,(-L_0-x)/\pi_2]$ any newly arriving demand is lost. Hence we have for the cumulative backlog at time $t=(-L_0-x)/\pi_2$,

$$c^I((-L_0-x)/\pi_2) = \frac{(-L_0-x)^2}{2\pi_2^2} + L_0\,\frac{(-L_0-x)}{\pi_2}.$$

At time $t=(-L_0-x)/\pi_2$ the inventory level equals $-L_0$. Using the lack of memory of the Poisson arrival process we find

$$c^I(x;M) = \frac{(-L_0-x)^2}{2\pi_2} + \frac{L_0(-L_0-x)}{\pi_2} + c^I(-L_0;M), \quad x < -L_0,$$

which equation is identical to (4.2.12).

Using equations (4.2.1)-(4.2.12) in the right order we find tractable expressions for basic functions associated with the models studied in chapter 1, 2 and 3.

Remark 4.2.1. Since each of the approximations given in chapter 1 and 2 is exact for exponentially distributed demand, and since the relations derived in chapter 3 and this chapter are exact, the resulting approximations for the present model are also exact for exponential demand.

Remark 4.2.2. In chis chapter we have considered a production-inventory model with complete lost-sales. In the paper by De Kok and Tijms [1985b] another production-inventory model with complete lost-sales is studied. There it is assumed that any demand that cannot be met directly from stock on hand is completely lost. It is not possible to derive approximations for this model using the analysis developed in the first four chapters of this

monograph. In their paper De Kok and Tijms [1985b] resort to two-moment approximations (cf. section 1.5).

Remark 4.2.3. If production rate $\pi_1=0$ and the inventory is controlled by an (m,M)-rule with M=m=0, the process $\{-X(t),\ t\geq 0\}$ corresponds to the workload process in an M/G/1-queue with impatient customers where the workload is processed at a constant rate of π_2 per unit time. The customer's impatience should be interpreted as follows. If a customer arrives at the queue and his waiting time exceeds a time τ then the customer leaves the system immediately. Using the results obtained in sections 4.1 and 4.2 approximations can be given for the fraction of work that is lost, the fraction of customers that is lost and the average workload; see De Kok and Tijms [1985c] for a more detailed study.

5. APPROXIMATIONS FOR THE AVERAGE HOLDING AND SWITCHING COSTS; THE OPTIMAL PRODUCTION QUANTITY.

In the previous chapters we focussed on deriving expressions for service measures. We motivated the use of service measures by the fact that in practice it is often hard to specify shortage costs. For a given (m,M)-rule we were able to give all wanted results. There it was assumed that the difference M-m was predetermined and the only goal was to find the unique level m such that the service level constraint was met. The choice of M-m was based on linear holding costs and a fixed switching cost. In section 1.5 we derived for M-m an EOQ-formula by using the deterministic version of the production-inventory model. Putting M-m equal to the EOQ-quantity we determined m. This sequential approach of determining M-m and m was motivated by the empirical finding that a good choice of M-m is typically rather insensitive to the required service level.

In this chapter we try to answer a few apparent questions:

1. Is it possible to find a tractable expression for the average holding and switching costs?
2. Does the (m,M)-rule, with M-m obtained from the EOQ-formula, show a good performance compared with the optimal (m,M)-rule with respect to average costs?
3. Is the optimal value of M-m indeed insensitive to the required service level?

We will show that all questions are more or less answered in the affirmative.

This chapter is organized as follows. In section 5.1 we express the total holding cost incurred during a cycle in terms of two basic functions. In section 5.2 we deal with the holding cost function corresponding to a production rate π_1 equal to 0, while in section 5.3 we derive expressions for the holding cost function for the cases of $\pi_1>0$ and $\pi_1<0$. The holding cost function corresponding to production rate π_2 is dealt with in section 5.4. In section 5.5 we give an approximation for the average holding and switching costs incurred per unit time. This result enables us to determine an approximate (m,M)-rule, which minimizes the average holding and switching costs per unit time subject to the requirement of a prespecified level for some service measures. Also, in section 5.5 it is argued that for the situation of high service the optimal value of M-m becomes independent of the specific level of the service measure considered. Moreover this constant value of M-m is the same for each of

the service measures discussed in chapters 1-4.
In section 5.6 we present numerical results in support of the quality of the approximations and the insensitivity of M-m to the required service level. Also some attention is paid to the diffusion process approach studied in Vickson [1982], using shortage costs rather than a service level constraint. All random variables, functions and quantities, which are not defined explicitly below, are defined as in the previous chapters.

5.1. General results for the holding cost per cycle.

In this section we derive general expressions for the holding cost incurred during a regeneration cycle. We recall that a cost at rate h.x is incurred when the on-hand-inventory equals x>0. Again we assume that at epoch 0 a cycle starts, i.e. X(0)=M. We define

C_h := the holding cost incurred during (0,T],

where T>0 is the next epoch at which the production rate is switched from π_2 to π_1. From the theory of regenerative processes we know that

(5.1.1) the average holding cost per unit time = $\frac{E[C_h]}{E[T]}$.

In order to find a tractable expression for $E[C_h]$ we define new basic functions. Given that X(0)=x+m, x≧0, and production rate π_1 is used, let

$k_1(x,m)$:= the expected holding cost incurred until the inventory level drops below m.

Given that X(0)=x≦M and production rate π_2 is used we define

$k_2(x)$:= the expected holding cost incurred until the inventory reaches the value M.

Then it immediately follows that under a given (m,M)-rule

(5.1.2) $E[C_h] = k_1(M-m,m) + \int_0^\infty k_2(m-u)\,d_u(1-p(M-m,u))$.

Next we reduce the problem of deriving an expression for $k_1(x,m)$ to the problem of finding an expression for $k_1(x,0)$. To do so imagine that the initial inventory x+m is divided into two amounts x and m and the amount m is separately kept in inventory during the time $T_1(x)$ needed to reduce the initial inventory level with at least x. Recalling that $t_1(x)=E[T_1(x)]$, we incur an expected holding cost $hmt_1(x)$ for the separate amount m, while an expected holding cost of $k_1(x,0)$ is incurred for the other inventory during the time $T_1(x)$. Thus

$$k_1(x,m) = k_1(x,0) + hmt_1(x), \quad x \geq 0. \tag{5.1.3}$$

Since we already derived an approximation for $t_1(x)$ we can restrict ourselves to finding an expression for $k_1(x,0)$. For ease of notation we define

$$k_1(x) := k_1(x,0), \quad x \geq 0.$$

It follows from (5.1.2) and (5.1.3) that

$$E[C_h] = k_1(M-m)+hmt_1(M-m) + \int_0^{\infty} k_2(m-u)d_u(1-p(M-m,u)). \tag{5.1.4}$$

Note that by their definitions the functions $t_1(x)$ and $k_1(x)$ are in fact strategy-independent.

So far we have not properly defined the model for which we want to derive expressions for the holding costs. It turns out that for any (m,M)-rule $E[C_h]$ is independent of the behaviour of arriving customers as long as the following two conditions are satisfied:

(i) $D_A \leq X(t_A^-) \Rightarrow X(t_A) = X(t_A^-)-D_A$,

(ii) $D_A > X(t_A^-) \Rightarrow X(t_A) \leq 0$,

where t_A is an arrival epoch and D_A is the demand of the arriving customer. Thus $E[C_h]$ *is the same for all models defined in chapters 1-4.* We will not rigorously prove this statement. The proof is based on the lack of memory of the exponential interarrival time distribution. Also, it follows that $k_2(x)$, $x \leq M$ *is independent of the customers behaviour,* provided conditions (i) and (ii) are satisfied. Since this also holds for $k_1(x)$, $t_1(x)$ and

$p(x,u)$ the above statement for $E[C_h]$ is in agreement with (5.1.4). Thus, in what follows it is no restriction to assume that *excess demand is backlogged.*

We conclude this section by giving expressions for $E[U^2]$. When M=m then an exact expression for $E[U^2]$ is computed from equation (1.3.34). When M-m is sufficiently large then we can compute an approximation for $E[U^2]$ from approximation (1.3.2).

Case M=m:

$$(5.1.5) \qquad E[U^2] = \begin{cases} 0 & \text{when } \pi_1 < 0 \\ E[D^2] & \text{when } \pi_1 = 0. \\ \frac{2\lambda}{\pi_1}\left(\frac{E[D^2]}{2s^*} - \frac{E[D]}{(s^*)^2} + \frac{\pi_1}{\lambda(s^*)^2}\right) & \text{when } \pi_1 > 0 \end{cases}$$

Case M>m: If M-m satisfies condition 1.3.1 then we have

Approximation 5.1.1.

$$E[U^2] \cong \begin{cases} \frac{\lambda E[D^3]}{3(\lambda E[D]-\pi_1)} & \text{when } \pi_1 \leq 0 \\ \frac{\lambda E[D^3]}{3(\lambda E[D]-\pi_1)} - \frac{2\lambda}{\lambda E[D]-\pi_1}\left(\frac{E[D^2]}{2s^*} - \frac{E[D]}{(s^*)^2} + \frac{\pi_1}{\lambda(s^*)^2}\right) & \text{when } \pi_1 > 0. \end{cases}$$

5.2. The function $k_1(x)$ for $\pi_1=0$.

In this section we consider the case of a production rate $\pi_1=0$. We derive an exact expression for $k_1(x)$ in terms of $E[U(x)]$ and $E[U^2(x)]$, where $U(x)$ is defined in section 1.2. Throughout this section we assume that at epoch 0 the inventory equals $x \geq 0$ and production rate $\pi_1=0$ is used.

We first recapitulate some definitions given in section 1.3. We define

τ_1 := the first arrival epoch after 0.

τ_n := the interarrival time between the (n-1)-th and n-th arrival, $n \geq 2$.

D_n := the demand of the n-th arriving customer, $n \geq 1$.

Letting

$$X_n := D_n - \pi_1 \tau_n, \quad n \geq 1,$$

define

$$S_0 := 0, \quad S_n := \sum_{j=1}^{n} X_j, \quad n \geq 1.$$

$$\zeta_0 := 0, \quad \zeta_k := \min\{n | n>\zeta_{k-1}, S_n > S_{\zeta_{k-1}}\}, \quad k \geq 1.$$

$$Z_k := S_{\zeta_k}, \quad k \geq 0.$$

$$N^*(x) := \min\{k | Z_k > x\}, \quad x \geq 0.$$

We refer to section 1.3 for an interpretation of these random variables. The above random variables play also a key role in the next section in which we derive expressions for $k_1(x)$ when $\pi_1 \neq 0$.

The random variable U(x) represents the undershoot of level 0, given that at epoch 0 the inventory equals x and production rate π_1 is used. For the present case of $\pi_1=0$ it follows that

$$U(x) = Z_{N^*(x)} - x, \quad x \geq 0.$$

Using the fact that for the case of $\pi_1=0$,

$$X_n = D_n, \quad n \geq 1,$$

$$Z_k = S_k, \quad k \geq 0,$$

it follows that

$$U(x) = \sum_{n=1}^{N^*(x)} D_n - x, \quad x \geq 0. \tag{5.2.1}$$

Next we derive an expression for $k_1(x)$. We define the random variable H(x) by

H(x) := the holding cost incurred until the inventory level decreases below 0

and hence

$$k_1(x) = E[H(x)], \quad x \geq 0.$$

We note that the inventory level is constant between arrival epochs. Then it is easy to verify that

$$H(x) = h\{x \sum_{n=1}^{N^*(x)} \tau_n - \sum_{n=1}^{N^*(x)} \tau_n \sum_{j=1}^{n-1} D_j\}.$$

Taking expectations and using the fact that $N^*(x)$ is a stopping time for $\{\tau_n\}$, we obtain

$$k_1(x) = h\{\frac{x}{\lambda} E[N^*(x)] - \frac{1}{\lambda} E[\sum_{n=1}^{N^*(x)} \sum_{j=1}^{n-1} D_j]\}. \tag{5.2.2}$$

It follows from (5.2.1) that

$$U^2(x) = \sum_{n=1}^{N^*(x)} D_n^2 + 2 \sum_{n=1}^{N^*(x)} D_n \sum_{j=1}^{n-1} D_j - 2x \sum_{n=1}^{N^*(x)} D_n + x^2.$$

Taking expectations and using the fact that $N^*(x)$ is also a stopping time for $\{D_n\}$, we find

$$E[U^2(x)] = E[N^*(x)]E[D^2] + 2E[D]E[\sum_{n=1}^{N^*(x)} \sum_{j=1}^{n-1} D_j] - 2xE[N^*(x)]E[D] + x^2. \tag{5.2.3}$$

A combination of the equations (5.2.2) and (5.2.3) then yields

$$k_1(x) = \frac{h}{\lambda E[D]} \{\frac{x^2}{2} - \frac{E[U^2(x)]}{2} + E[N^*(x)] \frac{E[D^2]}{2}\}. \tag{5.2.4}$$

We use again equation (5.2.1) to obtain

$$E[U(x)] = E[N^*(x)]E[D] - x. \tag{5.2.5}$$

Then (5.2.4) and (5.2.5) together imply

$$k_1(x) = \frac{h}{\lambda E[D]} \{\frac{x^2}{2} - \frac{E[U^2(x)]}{2} + \frac{E[D^2]}{2E[D]} (x + E[U(x)])\}. \tag{5.2.6}$$

It is important to point out that the lack of memory of the Poisson arrival process was only needed in order to have that τ_1 is distributed as τ_n, $n\geq 2$. Hence, in case $\{\tau_n\}$ is any sequence of independent and identically distributed random variables, we have

$$(5.2.7) \qquad k_1(x) = \frac{h\, E[\tau_1]}{E[D]} \{\frac{x^2}{2} - \frac{E[U^2(x)]}{2} + \frac{E[D^2]}{2E[D]} (x + E[U(x)])\}.$$

In the next section equation (5.2.7) will be applied to appropriately chosen random variables τ_n and D_n, $n\geq 1$, together with the corresponding random variable $U(x)$, $x\geq 0$. These random variables arise in a natural way when analyzing the two cases of $\pi_1>0$ and $\pi_1<0$ by random walk methods.

5.3. The function $k_1(x)$ for $\pi_1\neq 0$.

In this section we handle the cases $\pi_1>0$ and $\pi_1<0$ separately. The analysis is similar to that in section 1.3 and uses results there. In addition to the definitions given in section 5.2 we define

$$\nu_0 := 0, \quad \nu_k := \sum_{j=1}^{\zeta_k} \tau_j, \quad k \geq 1.$$

In section 1.3 we already introduced the random variables used in section 5.2 and interpreted these random variables in terms of the inventory process. Using these interpretations we observe that ν_k is the k-th arrival epoch after which the inventory level is strictly below the lowest inventory level that has been attained at arrival epochs up to epoch ν_k.

Throughout the remainder of this section we assume that at epoch 0 the starting inventory level $X(0)=x$ with $x\geq 0$, while production rate π_1 is *always* used. It follows from the definition of $H(x)$ given in section 5.2 that

$$(5.3.1) \qquad H(x) = h \int_0^{T_1(x)} X(t)dt, \quad x \geq 0,$$

where $T_1(x)$ is the time until the inventory level decreases below 0. We want to find an expression for $k_1(x)=E[H(x)]$. First we consider the case of $\pi_1>0$.

Expression for $k_1(x)$ when $\pi_1>0$.

For the case of $\pi_1>0$ it can be seen that

$$T_1(x) = \nu_{N^*(x)}, \quad x \geq 0.$$

We split up the interval $[0,T_1(x))$ by means of the intervals $\{[\nu_{k-1},\nu_k)\}_{k=1}^{N^*(x)}$. Then equation (5.3.1) implies that

$$H(x) = h\{\sum_{k=1}^{N^*(x)} \int_{\nu_{k-1}}^{\nu_k} X(\nu_{k-1})dt + \sum_{k=1}^{N^*(x)} \int_{\nu_{k-1}}^{\nu_k} (X(t)-X(\nu_{k-1}))dt\}, \quad x \geq 0. \tag{5.3.2}$$

We define

$$A_k := h \int_{\nu_{k-1}}^{\nu_k} (X(t)-X(\nu_{k-1}))dt, \quad k = 1,\ldots,N^*(x).$$

$$B(x) := h \sum_{k=1}^{N^*(x)} \int_{\nu_{k-1}}^{\nu_k} X(\nu_{k-1})dt, \quad x \geq 0.$$

Hence

$$H(x) = B(x) + \sum_{k=1}^{N^*(x)} A_k, \quad x \geq 0. \tag{5.3.3}$$

Some reflections show that B(x) can be interpreted as the holding cost incurred in the following demand model. Demands for a product occur at epochs generated by an arrival process $\{\nu_k\}$ and the k-th demand is distributed as Z_k-Z_{k-1}. No replenishments of inventory occur so that the inventory remains constant between the consecutive arrival epochs. Assuming that a holding cost is incurred at a rate being proportional to the on-hand inventory, we have that B(x) is the holding cost incurred until depletion of the initial inventory $x\geq 0$. Hence we have essentially the same model as studied in section 5.2 with the modifications (a) the arrival process of customers is given by a renewal process $\{\nu_k\}$ rather than by a Poisson process and (b) the demand of a customer is distributed as Z_1 rather than D_1. Here we use the property that $\{Z_k-Z_{k-1}\}$ and $\{\nu_k-\nu_{k-1}\}$ are independent sequences of independent and identically distributed random variables, as follows from the fact that the original process generating $\{\nu_k\}$ and $\{Z_k\}$ is a compound Poisson process. Thus we obtain an expression for B(x) by

replacing τ_1 and D by ν_1 and Z_1 in (5.2.7). As in the case of $\pi_1=0$, we also use that $U(x)=Z_{N^*(x)}-x$. This yields

$$E[B(x)] = \frac{hE[\nu_1]}{E[Z_1]}\{\frac{x^2}{2} - \frac{E[U^2(x)]}{2} + \frac{E[Z_1^2]}{2E[Z_1]}(x+E[U(x)])\}. \tag{5.3.4}$$

It can be seen that ν_1 is distributed as $T_1(0)$. Using the fact that $t_1(0)=E[T_1(0)]$ by definition, we obtain

$$E[\nu_1] = t_1(0).$$

From (1.3.33) we know that $t_1(0)=1/(\pi_1 s^*)$, where s^* is defined by (1.3.21). Hence we have that

$$E[\nu_1] = 1/(\pi_1 s^*). \tag{5.3.5}$$

We now focus on the random variables A_k, $1\leq k\leq N^*(x)$. Again, since the process underlying these variables is a compound Poisson process, it follows that the random variables A_k, $1\leq k\leq N^*(x)$ are independent and identically distributed random variables. Then it follows that

$$E[\sum_{n=1}^{N^*(x)} A_k] = E[N^*(x)]E[A_1]. \tag{5.3.6}$$

To find an expression for $E[A_1]$ we proceed along the same lines as in the derivation of an expression for $k_1(x)$ in section 5.2.

It follows from the definitions of τ_n, X_n, A_1 and ζ_1 that

$$A_1 = h\{\sum_{n=1}^{\zeta_1} \pi_1 \frac{\tau_n^2}{2} - \sum_{n=1}^{\zeta_1} \tau_n \sum_{j=1}^{n-1} X_j\}.$$

Because ζ_1 is a stopping time for $\{\tau_n\}$ we find

$$E[A_1] = h\{\frac{\pi_1 E[\zeta_1]}{\lambda^2} - \frac{1}{\lambda} E[\sum_{n=1}^{\zeta_1} \sum_{j=1}^{n-1} X_j]\}. \tag{5.3.7}$$

We also have that

$$Z_1 = \sum_{n=1}^{\zeta_1} X_j$$

and hence

$$(5.3.8) \qquad E[Z_1] = E[\zeta_1]E[X_1]$$

and

$$(5.3.9) \qquad E[Z_1^2] = E[\zeta_1]E[X_1^2] + 2E[X_1]E[\sum_{n=1}^{\zeta_1}\sum_{j=1}^{n-1} X_j],$$

where we used that ζ_1 is a stopping time for $\{X_n\}$. Combining the equations (5.3.7)-(5.3.9) and using the expressions for $E[Z_1]$ and $E[Z_1^2]$ given by (1.3.22) and (1.3.23), we obtain

$$(5.3.10) \qquad E[A_1] = \frac{h}{\pi_1 (s^*)^2} .$$

Now we are in a position to give an exact expression for $k_1(x)$. The fact that $U(x) = Z_{N^*(x)} - x$ implies

$$U(x) = \sum_{n=1}^{N^*(x)} (Z_k - Z_{k-1}) - x$$

and by an application of Wald's equation we thus find

$$(5.3.11) \qquad E[N^*(x)] = \frac{x+E[U(x)]}{E[Z_1]} .$$

Then the equations (5.3.3)-(5.3.6), (5.3.10) and (5.3.11) together yield after some algebra

$$(5.3.12) \qquad k_1(x) = \frac{h}{\lambda E[D]-\pi_1}\{\frac{x^2}{2} - \frac{E[U^2(x)]}{2} + \frac{\lambda E[D^2]}{2(\lambda E[D]-\pi_1)}(x+E[U(x)])\}.$$

Again we used the expressions for $E[Z_1]$ and $E[Z_1^2]$ given in section 1.3.

This concludes the derivation of an exact expression for $k_1(x)$ for the case of $\pi_1 > 0$. Next we derive an exact expression for $k_1(x)$ for the case of $\pi_1 < 0$.

Expression for $k_1(x)$ when $\pi_1 < 0$.

For the case of $\pi_1 < 0$ we proceed along the same lines as for the case of $\pi_1 > 0$. We note that

$$\nu_k = \sum_{j=1}^{k} \tau_k, \qquad k \geq 0.$$

$$Z_k = S_k, \quad k \geq 0.$$

We define for x>0

R(x) := the time that elapses between the downcrossing of 0 and the next arrival, when the inital inventory is x.

We first derive an expression for H(x). To do so we consider realizations of the inventory process during the time interval $[0,\nu_{N^*(x)})$. Let ω be an element of the underlying sample space Ω. For any random variable X define

X_ω := the realization of X corresponding to ω.

Next we split up Ω into two disjoint sets,

$$W_0 := \{\omega | R_\omega(x)=0\}$$

$$W_+ := \{\omega | R_\omega(x)>0\}.$$

We first consider the case of $\omega \in W_0$. If $R_\omega(x)=0$ then the level 0 is downcrossed for the first time through the arrival of a customer. The reader may find it helpful to draw a picture to verify the relation

$$H_\omega(x) = h \sum_{k=1}^{N^*_\omega(x)} X_\omega(\nu_{k-1,\omega})\tau_{k,\omega} - h(-\pi_1) \sum_{k=1}^{N^*_\omega(x)} \frac{\tau^2_{k,\omega}}{2}, \quad \omega \in W_0. \tag{5.3.13}$$

Next we consider the case of $\omega \in W_+$. If $R_\omega(x)>0$ then the level 0 is downcrossed during the time interval $(\nu_{N^*_\omega(x)-1,\omega}, \nu_{N^*_\omega(x),\omega})$. Again a picture would help to see that

$$H_\omega(x) = h \sum_{k=1}^{N^*_\omega(x)-1} X_\omega(\nu_{k-1,\omega})\tau_{k,\omega} - h(-\pi_1) \sum_{k=1}^{N^*_\omega(x)-1} \frac{\tau^2_{k,\omega}}{2} \tag{5.3.14}$$

$$+ h(-\pi_1)\tfrac{1}{2}(\tau_{N^*_\omega(x),\omega} - R_\omega(x))^2, \quad \omega \in W_+.$$

Some reflection shows that

$$R_\omega(x) = \tau_{N^*_\omega(x),\omega} - (x - S_{N^*_\omega(x)-1,\omega})/(-\pi_1), \quad \omega \in W_+. \tag{5.3.15}$$

Using $x-S_{N^*(x)-1} = X(\nu_{N^*(x)-1})$, we find by substitution of the equation (5.3.15) into (5.3.14) that

$$(5.3.16)\qquad H_\omega(x) = h \sum_{k=1}^{N^*_\omega(x)} X_\omega(\nu_{k-1,\omega})\tau_{k,\omega} - h(-\pi_1) \sum_{k=1}^{N^*_\omega(x)} \frac{\tau^2_{k,\omega}}{2} + h(-\pi_1)\frac{R^2_\omega(x)}{2}, \qquad \omega\in\bar{W}_+.$$

Combining (5.3.13) and (5.3.16) we obtain

$$(5.3.17)\qquad H(x) = h \sum_{k=1}^{N^*(x)} X(\nu_{k-1})\tau_k - h(-\pi_1) \sum_{k=1}^{N^*(x)} \frac{\tau^2_k}{2} + h(-\pi_1)\frac{R^2(x)}{2}.$$

We again define

$$B(x) := h \sum_{k=1}^{N^*(x)} X(\nu_{k-1})\tau_k.$$

Now B(x) can be interpreted as the holding cost incurred until the depletion of an initial inventory x in a pure demand model in which demands occur at epochs generated by a Poisson process with rate λ. Each demand is distributed as the generic random variable

$$Z_1 := D + (-\pi_1)\tau_1.$$

The demands are independent of the arrival process. Similarly to the derivation of (5.3.4) we now use (5.2.7). However, there is an important difference with the case of $\pi_1 \geq 0$. Because the level 0 can be downcrossed between arrival epochs we have no longer that U(x) equals $Z_{N^*(x)}-x$. Therefore we define the random variable $\tilde{U}(x)$ by

$$\tilde{U}(x) := Z_{N^*(x)} - x.$$

Then it follows from equation (5.2.7) with $E[\nu_1]=1/\lambda$, D replaced by Z_1 and U(x) replaced by $\tilde{U}(x)$ that

$$(5.3.18)\qquad E[B(x)] = \frac{h}{\lambda E[Z_1]}\{\frac{x^2}{2} - \frac{E[\tilde{U}^2(x)]}{2} + \frac{E[Z_1^2]}{2E[Z_1]}(x+E[\tilde{U}(x)])\}.$$

Applying Wald's equation we obtain

$$(5.3.19)\qquad E\Big[\sum_{k=1}^{N^*(x)} \tau_k^2\Big] = \frac{2}{\lambda^2} E[N^*(x)].$$

Analogously to (5.3.11) we find from the definition of $\tilde{U}(x)$ that

$$(5.3.20)\qquad E[N^*(x)] = \frac{x+E[\tilde{U}(x)]}{E[Z_1]}.$$

So it remains to find expressions for $E[\tilde{U}(x)]$, $E[\tilde{U}^2(x)]$ and $E[R^2(x)]$.

We now express U(x) in terms of $\tilde{U}(x)$ and R(x). It follows from the definitions of these three random variables that

$$\tilde{U}(x) = \begin{cases} U(x) & \text{when } U(x) > 0 \\ (-\pi_1)R(x)+D_{N^*(x)} & \text{when } U(x)=0 \end{cases}$$

or equivalently

$$(5.3.21)\qquad \tilde{U}(x) = U(x) + ((-\pi_1)R(x) + D_{N^*(x)})1_{\{U(x)=0\}}.$$

Then it is easy to see that

$$(5.3.22)\qquad \tilde{U}^2(x) = U^2(x) + ((-\pi_1)R(x) + D_{N^*(x)})^2 1_{\{U(x)=0\}}.$$

We now make the crucial observation that, *given* U(x)=0, R(x) and $D_{N^*(x)}$ are independent random variables, and moreover, R(x) is exponentially distributed with mean $1/\lambda$ while $D_{N^*(x)}$ has distribution function F. This follows from the lack of memory of the Poisson arrival process and the fact that, *given* U(x)=0, the first demand occurring after the downcrossing of 0 must be $D_{N^*(x)}$. Thus we obtain from these observations and the equations (5.3.21) and (5.3.22) that

$$(5.3.23)\qquad E[\tilde{U}(x)] = E[U(x)] + (E[D]-\pi_1/\lambda)P\{U(x)=0\}$$

$$(5.3.24)\qquad E[\tilde{U}^2(x)] = E[U^2(x)] + (E[D^2]- \frac{2\pi_1}{\lambda} (E[D]- \frac{\pi_1}{\lambda}))P\{U(x)=0\}$$

In a similar fashion we obtain

(5.3.25) $$E[R^2(x)] = \frac{2}{\lambda^2} P\{U(x)=0\}.$$

To find an expression for $k_1(x)$ we substitute (5.3.18)-(5.3.20) and (5.3.23)-(5.3.25) into equation (5.3.17). Some straightforward algebra then yields

(5.3.26) $$k_1(x) = \frac{h}{\lambda E[D]-\pi_1}\{\frac{x^2}{2} - \frac{E[U^2(x)]}{2} + \frac{\lambda E[D^2]}{2(\lambda E[D]-\pi_1)}(x+E[U(x)])\}.$$

Comparing this result with (5.2.6) and (5.3.12) shows that for each of the three cases $\pi_1=0$, $\pi_1>0$ and $\pi_1<0$, $k_1(x)$ is given by the same expression.

5.4. The function $k_2(x)$.

In this section we assume that at epoch 0 the inventory level $X(0)=x\leqq M$ and that the production is always governed by rate π_2. We recall the definition of $T_2(x)$ given in section 1.4,

$T_2(x)$:= the time until the inventory level reaches the value M, $x \leqq M$.

Then it follows from the definition of $k_2(x)$ that

(5.4.1) $$k_2(x) = hE[\int_0^{T_2(x)} X(t)1_{\{X(t)>0\}}dt].$$

Also it is obvious that

(5.4.2) $$E[\int_0^{T_2(x)} X(t)1_{\{X(t)>0\}}dt] = E[\int_0^{T_2(x)} X(t)dt] + E[\int_0^{T_2(x)} -X(t)1_{\{X(t)<0\}}dt].$$

But from the definition of c(x) given in section 1.2 we have that

(5.4.3) $$c(x) = E[\int_0^{T_2(x)} -X(t)1_{\{X(t)<0\}}dt].$$

We further note that

(5.4.4) $$E[\int_0^{T_2(x)} X(t)dt] = E[\int_0^{T_2(x)} Mdt] - E[\int_0^{T_2(x)} (M-X(t))dt].$$

Combining (5.4.1)-(5.4.4) and using $t_2(x)=E[T_2(x)]$ we obtain

$$(5.4.5) \qquad k_2(x) = h\{Mt_2(x)-E[\int_0^{T_2(x)} (M-X(t))dt] + c(x)\}$$

It remains to determine $E[\int_0^{T_2(x)} (M-X(t))dt]$. In section 1.4 we derived equation (1.4.36), which gives an expression for $c(x)$, $x\leq 0$. The arguments used there can also be applied here. The process $\{M-X(t),\ 0\leq t\leq T_2(x)\}$ corresponds to the workload process in an M/G/1-queue in which jobs arrive according to a Poisson process with rate λ. The amounts of work involved by the jobs are independent random variables with common distribution function F and work is processed at rate π_2. The initial workload equals M-x. Using again the results in Tijms [1977] we find that

$$(5.4.6) \qquad E[\int_0^{T_2(x)} (M-X(t))dt] = \frac{(M-x)^2}{2(\pi_2-\lambda E[D])} + \frac{\lambda E[D^2](M-x)}{2(\pi_2-\lambda E[D])^2}, \quad x \leq M.$$

Thus we can give the following expression for $k_2(x)$, $x\leq M$ from equations (1.4.1), (5.4.5) and (5.4.6),

$$(5.4.7) \qquad k_2(x) = h\{\frac{M^2-x^2}{2(\pi_2-\lambda E[D])} - \frac{\lambda E[D^2](M-x)}{2(\pi_2-\lambda E[D])^2} + c(x)\}, \quad x \leq M.$$

Hence we have expressed $k_2(x)$ in terms of $c(x)$ for which function we already found approximation 1.4.3. Note that it follows from equations (1.4.36) and (5.4.7) that

$$k_2(x) = k_2(0), \quad x \leq 0.$$

This also follows directly from the lack of memory of the exponential interarrival time distribution. Finally we derive the following result from equations (1.2.10), (5.4.7) and the definition of U(x),

$$(5.4.8) \qquad \int_0^{\infty} k_2(m-u)d_u(1-p(M-m,u)) = h\{\frac{(M-m)^2-E[U(M-m)]^2}{2(\pi_2-\lambda E[D])} + \frac{(m-\lambda E[D^2])}{2(\pi_2-\lambda E[D])} \frac{(M-m+E[U(M-m)])}{(\pi_2-\lambda E[D])} + E[C]\}.$$

In the next section we will combine the results given in sections 5.2-5.4 to give an expression for $E[C_h]$.

5.5. The average holding and switching costs per unit time and insensitivity results for M-m.

In sections 5.2-5.4 we have found exact expressions for $k_1(x)$ and $k_2(x)$. From these results we can give an exact expression for $E[C_h]$. From this exact result we arrive at a computationally tractable approximation for $E[C_h]$.

From equation (1.3.10) we have that

$$t_1(M-m) = \frac{M-m+E[U(M-m)]}{\lambda E[D]-\pi_1} . \tag{5.5.1}$$

It then follows from equations (5.1.4), (5.2.6), (5.3.12), (5.3.26), (5.4.8) and (5.5.1) that

$$\begin{aligned} E[C_h] = h\{&\frac{(\pi_2-\pi_1)}{(\lambda E[D]-\pi_1)(\pi_2-\lambda E[D])}\left[\frac{(M-m)^2}{2} - \frac{E[U^2(M-m)]}{2}\right. \\ &\left. + m(M-m+E[U(M-m)])\right] + E[C] \\ &+ \frac{\lambda E[D^2]}{2}(M-m+E[U(M-m)])\left[\frac{1}{(\lambda E[D]-\pi_1)^2} - \frac{1}{(\pi_2-\lambda E[D])^2}\right]\}. \end{aligned} \tag{5.5.2}$$

Now we recall that the random variable U was defined as

$$U := U(M-m).$$

For the case of M=m exact expressions for $E[U]$ and $E[U^2]$ are given by equations (1.3.35) and (5.1.5). If M-m satisfies condition 1.3.1 then approximations for $E[U]$ and $E[U^2]$ are given by approximations 1.3.3 and 5.1.1. Letting $E_{app}[C]$ denote the approximation for $E[C]$ resulting from equations (1.2.10) and (1.4.36) and approximation 1.4.3, (5.5.2) yields

Approximation 5.5.1. If M=m or M-m satisfies condition 1.3.1 then

$$\begin{aligned} E[C_h] \cong h\{&\frac{(\pi_2-\pi_1)}{(\lambda E[D]-\pi_1)(\pi_2-\lambda E[D])}\left[\frac{(M-m)^2}{2} - \frac{E[U^2]}{2} + m(M-m+E[U])\right] \\ &+ E_{app}[C] + \frac{\lambda E[D^2]}{2}(M-m+E[U])\left[\frac{1}{(\lambda E[D]-\pi_1)^2} - \frac{1}{(\pi_2-\lambda E[D])^2}\right]\}. \end{aligned}$$

We define

$K(\Delta,m)$:= the expected holding and switching costs incurred during a cycle in case the inventory is controlled by an $(m,m+\Delta)$-rule.

Under the assumption that *the customer's behaviour satisfies the two conditions given in section* 5.1, we obtain

Approximation 5.5.2. If $\Delta=0$ or Δ satisfies condition 1.3.1, then

$$K(\Delta,m) \cong K + h\{\frac{(\pi_2-\pi_1)}{(\lambda E[D]-\pi_1)(\pi_2-\lambda E[D])}[\frac{\Delta^2}{2} - \frac{E[U^2]}{2} + m(\Delta+E[U])] +$$

$$+ E_{app}[C] + \frac{\lambda E[D^2]}{2}(\Delta+E[U])[\frac{1}{(\lambda E[D]-\pi_1)^2} - \frac{1}{(\pi_2-\lambda E[D])^2}]\}.$$

Again we emphasize the fact that $K(\Delta,m)$ defined above is the same for all models defined in chapters 1-4. Let us consider a particular model. Denote by $E[T(\Delta,m)]$ and $E[T_B(\Delta)]$ the expected length of a cycle in the model under consideration and the pure backlog model, respectively. Define the random variable $B_2(\Delta,m)$ as in section 3.1,

$B_2(\Delta,m)$:= amount of demand lost during a cycle in the model under consideration.

Then it follows from arguments given in section 2.3 to derive equation (2.3.22) that

$$E[T(\Delta,m)] = E[T_B(\Delta)] - \frac{E[B_2(\Delta,m)]}{\pi_2-\lambda E[D]}.$$

From equations (1.3.10), (1.3.31), approximation 1.3.3 and equations (1.2.8) and (1.4.1), we obtain the following approximation,

Approximation 5.5.3. If $\Delta=0$ or Δ satisfies condition 1.3.1, then

$$E[T(\Delta,m)] \cong \frac{(\pi_2-\pi_1)(\Delta+E[U])}{(\lambda E[D]-\pi_1)(\pi_2-\lambda E[D])} - \frac{E[B_2(\Delta,m)]}{(\pi_2-\lambda E[D])}.$$

Define

$g(\Delta,m)$:= the long-run average cost per unit time.

Since by the theory of regenerative processes,

$$g(\Delta,m) = \frac{K(\Delta,m)}{E[T(\Delta,m)]}$$

we find from approximations 5.5.2 and 5.5.3 and the results given in chapter 1-4 an approximation for $g(\Delta,m)$ associated with the particular model under consideration.

Now we are able to find an approximately average-cost optimal $(m,m+\Delta)$-rule such that a prespecified level of some service measure is met. Define for a given service measure

$$r(\Delta,m) := \text{the service level under an } (m,m+\Delta)\text{-rule.}$$

We are interested in the approximate solution to the following problem.

Problem Pb 1:
Find (Δ^*,m^*) such that

$$g(\Delta^*,m^*) = \min\{g(\Delta,m) \mid \Delta\geq 0,\ m\geq 0,\ r(\Delta,m)=\alpha\},$$

where α is the prespecified level of the service measure under consideration.

Problem Pb 1 can be solved as follows. For each $\Delta\geq 0$ there exists at most one $m(\Delta)$ such that $r(\Delta,m(\Delta))=\alpha$. Provided the service level α is sufficiently high there is a range of Δ-values for which the function $m(\Delta)$ is defined. For each Δ in this range $m(\Delta)$ can be determined by bisection, where we use the strict monotonicity of $r(\Delta,m)$ in m. Next we use a simple line search method to obtain Δ^* from

$$g(\Delta^*,m(\Delta^*)) = \min_{\Delta} g(\Delta,m(\Delta)).$$

Thus the two-dimensional problem is reduced to two one-dimensional problems, which is computationally favourable.

Our numerical investigations revealed the remarkable result that the value of the approximately optimal Δ^* became almost independent of the required service level α for α sufficiently close to 1, say $\alpha\geq 0.99$.

Moreover, it turned out that this constant value $\tilde{\Delta}^*$ (say) of Δ was the same for each of the service measures considered and for all models studied in chapter 1-4. We found these empirical findings to be surprising at first sight, but realized next that they could not be accidental. Indeed these findings suggest that the various service measures must have some asymptotic behaviour in common.

To make specific this common asymptotic behaviour let us consider the expected amount of demand backlogged during a cycle in the backlog model. This quantity is given by

$$(5.5.3)\qquad E[B(\Delta,m)] = \int_0^m b_\infty(m-u)\,d_u(1-p(\Delta,u)) + b_\infty(0)p(\Delta,m) + \frac{\pi_2}{(\pi_2-\lambda E[D])}\int_m^\infty p(\Delta,u)\,du - b_\infty(m+\Delta),$$

where $b_\infty(x)$ is defined in section 1.4. Note that we have made explicit the dependence of B on Δ and m. From equation (1.4.32) we know that

$$(5.5.4)\qquad \lim_{x\to\infty} e^{\delta x}\, b_\infty(x) = \frac{1}{\delta^2\nu}.$$

We show below that

$$(5.5.5)\qquad \lim_{m\to\infty} e^{\delta m}\, E[B(\Delta,m)] = c$$

for some c. Thus for m sufficiently large

$$E[B(\Delta,m)] \cong ce^{-\delta m}.$$

This result is the key to the understanding of the insensitivity of Δ to the required service level α when α is sufficiently close to 1. We make our point clear after proving (5.5.5).

Because of the distribution assumption we have that positive constants C_1, κ exist such that

$$(5.5.6)\qquad 1-F(x) \leq C_1 e^{-\kappa x} \quad \text{for all } x \geq 0.$$

It is no restriction to assume

$$(5.5.7)\qquad \kappa > \delta.$$

In section 1.3 we defined a renewal process $\{Z_k\}$, a random variable $N^*(x)$ and the associated renewal function $M^*(x)=E[N^*(x)]$. Equation (1.3.12) states that for $\pi_1 \geqq 0$

$$p(x,u) = P\{Z_{N^*(x)} - x > u\}, \quad x \geq 0, \ u \geq 0.$$

Conditioning on the last renewal before "epoch" x we find

$$p(x,u) = \int_0^x P\{Z_1 > u+x-y\} dM^*(y). \tag{5.5.8}$$

It follows from (1.3.15), (1.3.20) and (5.5.6) that for the case of $\pi_1 \geqq 0$

$$P\{Z_1 > u\} \leqq C_2 e^{-\kappa u}, \quad u \geq 0,$$

for some positive constant C_2. Substituting this result into (5.5.8) it follows that for the case of $\pi_1 \geqq 0$

$$p(x,u) \leqq C_2 M^*(x) e^{-\kappa u}, \quad u \geq 0, \ x \geq 0. \tag{5.5.9}$$

Using the arguments used to obtain (1.3.30) it can be seen that (5.5.9) also holds for the case of $\pi_1 < 0$. Using (5.5.7) and (5.5.9) we find for all values of π_1

$$\lim_{m\to\infty} e^{\delta m} p(\Delta,m) = 0. \tag{5.5.10}$$

$$\lim_{m\to\infty} e^{\delta m} \int_m^\infty p(\Delta,u)du = 0. \tag{5.5.11}$$

After some algebra we further find from (5.5.4) and (5.5.9) that

$$\lim_{m\to\infty} e^{\delta m} \int_0^m b_\infty(m-u) d_u(1-p(\Delta,u)) = \frac{1}{\delta^2\nu} \int_0^\infty e^{\delta u} d_u(1-p(\Delta,u)). \tag{5.5.12}$$

Since (5.5.4) implies that as $m\to\infty$ then $e^{\delta m} b_\infty(m+\Delta) \to e^{-\delta\Delta}/(\delta^2\nu)$, we finally obtain from (5.5.3) and (5.5.10)-(5.5.12)

$$\lim_{m\to\infty} e^{\delta m} E[B(\Delta,m)] = \frac{1}{\delta^2\nu} \left[\int_0^\infty e^{\delta u} d_u(1-p(\Delta,u)) - e^{-\delta\Delta} \right]. \tag{5.5.13}$$

Analogously we find for the expected cumulative backlog during a cycle in the backlog model,

$$(5.5.14)\qquad \lim_{m\to\infty} e^{\delta m}E[C(\Delta,m)] = \frac{1}{\pi_2\delta^3\nu}\left[\int_0^\infty e^{\delta u}d_u(1-p(\Delta,u))-e^{-\delta\Delta}\right].$$

In general we can state the following result. For each of the models studied in chapter 1-4 and each service measure considered there we have that

$$(5.5.15)\qquad \lim_{m\to\infty} e^{\delta m}(1-r(\Delta,m))E[T(\Delta,m)] = c_r\left[\int_0^\infty e^{\delta u}d_u(1-p(\Delta,u))-e^{-\delta\Delta}\right]$$

where $r(\Delta,m)$ is the level of the service measure considered and c_r is some positive constant also depending on the service measure dealt with. Equation (5.5.15) describes the common asymptotic behaviour that we mentioned earlier.

Now we apply the approximation for $p(\Delta,u)$ given in section 1.3. Using the definitions of δ and s^* for the particular case of $\pi_1>0$, we obtain for all values of π_1 and all Δ satisfying condition 1.3.1

$$(5.5.16)\qquad \int_0^\infty e^{\delta u}d_u(1-p(\Delta,u)) \cong (\pi_2-\pi_1)[(\lambda E[D]-\pi_1)(1+\delta(\frac{\lambda E[D^2]}{2(\lambda E[D]-\pi_1)}-E[U]))]^{-1}.$$

Substituting (5.5.14)-(5.5.16) into approximations 5.5.1 and 5.5.3, we obtain

Approximation 5.5.4. For m sufficiently large and Δ satisfying condition 1.3.1 we have,

$$g(\Delta,m) \cong \tilde{g}(\Delta,m)$$

with

$$\begin{aligned}\tilde{g}(\Delta,m) = [K+h\{&\frac{(\pi_2-\pi_1)}{(\lambda E[D]-\pi_1)(\pi_2-\lambda E[D])}\left(\frac{\Delta^2}{2}-\frac{E[U^2]}{2}+m(\Delta+E[U])\right)\\ &+\frac{1}{\pi_2\delta^3\nu}(c_\delta-e^{-\delta\Delta})e^{-\delta m}+\\ &+\frac{\lambda E[D^2]}{2}(\Delta+E[U])\left(\frac{1}{(\lambda E[D]-\pi_1)^2}-\frac{1}{(\pi_2-\lambda E[D])^2}\right)\}]\times\\ &\times\left[\frac{(\pi_2-\pi_1)(\Delta+E[U])}{(\lambda E[D]-\pi_1)(\pi_2-\lambda E[D])}-c_\ell(c_\delta-e^{-\delta\Delta})e^{-\delta m}/(\pi_2-\lambda E[D])\right]^{-1}\end{aligned}$$

and where

$$c_\delta := (\pi_2-\pi_1)[(\lambda E[D]-\pi_1)(1+\delta(\frac{\lambda E[D^2]}{2(\lambda E[D]-\pi_1)} - E[U]))]^{-1}$$

$$c_\ell := \lim_{m\to\infty} e^{\delta m} b_{2,\infty}(m),$$

where $b_{2,\infty}(x)$ is defined as

$b_{2,\infty}(x)$:= the expected amount of demand lost during the interval $(0,\infty)$ in the particular model under consideration, given that $X(0)=x$ and production rate π_2 is always used.

For the important pure backlog and pure lost-sales model we have

$$c_\ell = \begin{cases} 0 & \text{for the backlog model} \\ \dfrac{\pi_2-\lambda E[D]}{\pi_2\delta^2\nu} & \text{for the lost-sales model.} \end{cases}$$

Let us consider a service measure such that the associated level $r(\Delta,m)$ for an $(m,m+\Delta)$-rule satisfies (5.5.15). Define

$$\tilde{r}(\Delta,m) := 1-c_r(c_\delta-e^{-\delta\Delta})e^{-\delta m}[\frac{(\pi_2-\pi_1)(\Delta+E[U])}{(\lambda E[D]-\pi_1)(\pi_2-\lambda E[D])}-c_\ell(c_\delta-e^{-\delta\Delta})e^{-\delta m}]^{-1}.$$

where c_r is given through (5.5.15). Then we solve the following new problem being an approximation of problem Pb 1,

Problem $\tilde{Pb}$ 1:
Find $(\tilde{\Delta}^*,\tilde{m}^*)$ such that

$$\tilde{g}(\tilde{\Delta}^*,\tilde{m}^*) = \min\{\tilde{g}(\Delta,m) \mid \Delta\geq 0,\ m\geq 0,\ \tilde{r}(\Delta,m)=\alpha\}$$

where α is the prespecified level of the service measure under consideration.

If both $\tilde{\Delta}^*>0$ and $\tilde{m}^*>0$ then it follows from the Lagrange multiplier method that there exists an $\tilde{\eta}^*$ such that

$$\frac{\partial}{\partial \Delta}\left[\tilde{g}(\Delta,m) - \eta(\tilde{r}(\Delta,m)-\alpha)\right]_{(\Delta,m,\eta)=(\tilde{\Delta}^*,\tilde{m}^*,\tilde{\eta}^*)} = 0$$

$$\frac{\partial}{\partial m}\left[\tilde{g}(\Delta,m) - \eta(\tilde{r}(\Delta,m)-\alpha)\right]_{(\Delta,m,\eta)=(\tilde{\Delta}^*,\tilde{m}^*,\tilde{\eta}^*)} = 0$$

$$\tilde{r}(\tilde{\Delta}^*,\tilde{m}^*) = \alpha.$$

Solving this set of equations we find after a lot of tedious algebra that $\tilde{\Delta}^*$ is a positive root of

$$z(\Delta) = 0,$$

where

$$(5.5.17) \qquad z(\Delta) := h\Delta(\Delta+E[U]) + \frac{h(\Delta+E[U])^2 e^{-\delta\Delta}}{c_\delta - e^{-\delta\Delta}} - \frac{h(\Delta^2-E[U^2])}{2}$$

$$- \frac{h(\Delta+E[U])}{\delta} - K\frac{(\lambda E[D]-\pi_1)(\pi_2-\lambda E[D])}{(\pi_2-\pi_1)}.$$

We do not have analytical results concerning the existence and uniqueness of $\tilde{\Delta}^*$. In all numerical examples tested we found a unique positive solution. Also, the value of $\tilde{\Delta}^*$ we obtained from solving $z(\Delta)=0$ is practically equal to the value of Δ^* we found by solving Pb 1 when α is sufficiently close to 1. We observe that the expression for $z(\Delta)$ involves neither the model constant c_ℓ nor the service measure constant c_r. Hence we have found the following important result.

Insensitivity result: The optimal value $\tilde{\Delta}^*$ that is found by solving $\tilde{Pb}$ 1 is independent of the model and service measure under consideration.

Now we suggest the following approximate solution to Pb 1 when α is sufficiently close to 1 (say, $\alpha \geq 0.99$).

1. Compute $\tilde{\Delta}^*$ from $z(\tilde{\Delta})^*=0$ with $z(\Delta)$ given by (5.5.17).
2. Compute $m^* \geq 0$ from

$$r(\tilde{\Delta}^*,m^*) = \alpha.$$

Step 2 can be solved e.g. by bisection, where we use the fact that $r(\Delta,m)$ is strictly increasing in m for fixed Δ. Step 1 can be solved by the standard Newton-Raphson method with either exact or approximate derivative. We suggest to start this iterative procedure with

$$(5.5.18) \qquad \Delta_0 = \frac{1}{\delta} - E[U] + \{(E[U])^2 + \frac{1}{\delta^2} - E[U^2] + \frac{2K(\lambda E[D]-\pi_1)(\pi_2-\lambda E[D])}{h(\pi_2-\pi_1)}\}^{\frac{1}{2}}.$$

This suggestion is motivated as follows. For all $\Delta \geq 0$ the function $z(\Delta)$ satisfies the inequality $z(\Delta)>w(\Delta)$ with

$$w(\Delta) := h\{\Delta(\Delta+E[U]) - \frac{(\Delta^2-E[U^2])}{2} - \frac{(\Delta+E[U])}{\delta}\} - K\frac{(\lambda E[D]-\pi_1)(\pi_2-\lambda E[D])}{(\pi_2-\pi_1)}.$$

Then we have that Δ_0 is the largest root of $w(\Delta)=0$. If there exists a positive solution $\tilde{\Delta}^*$ of (5.5.18) then $w(\Delta)=0$ has two real roots and $\tilde{\Delta}^*<\Delta_0$.

In the next section we will give numerical results concerning the accuracy of the approximation 5.5.1 for the average holding costs per cycle. Also, we will plot the approximately optimal values (Δ^*,m^*) obtained from the algorithm described below problem Pb 1 and show the convergence of Δ^* to $\tilde{\Delta}^*$ as α increases.

Remark 5.5.1. Consider the case of exponentially distributed demand. Then

$$\tilde{g}(\Delta,m) = g(\Delta,m) \text{ and } \tilde{r}(\Delta,m) = r(\Delta,m).$$

Also $g(\Delta,m)$ and $r(\Delta,m)$ are exact, provided $r(\Delta,m)$ corresponds to one of the service measures discussed before. Hence for the exponential demand case the solution $(\tilde{\Delta}^*,\tilde{m}^*)$ is the true optimal solution.

Remark 5.5.2. Consider the backlog model in which shortage costs are assumed rather than a service level constraint. Suppose that a shortage cost at rate p.x is incurred whenever a backlog of x exists. Then, similarly as above, we can approximately solve the problem of minimizing the average costs, where the costs consist of holding, switching and shortage costs. The average shortage cost equals p times the average backlog at an arbitrary point in time, the latter being studied in chapter 1. The $(m^*,m^*+\Delta^*)$-rule that minimizes the average costs has again the property

that $\Delta^* \cong \tilde{\Delta}^*$, especially when the shortage cost rate p is large, causing m^* to be large.

Let us define

$v(\Delta,m)$:= the sum of the average holding, switching and shortage costs when the inventory is controlled by an $(m,m+\Delta)$-rule

and consider the problem

Problem Pb 2:
Find (Δ^*,m^*) such that

$$v(\Delta^*,m^*) = \min\{v(\Delta,m) \mid \Delta,m\geq 0\}.$$

We suggest to solve approximately this problem by the following procedure.
1. Compute $\tilde{\Delta}^*$ from $z(\tilde{\Delta}^*)=0$ with $z(\Delta)$ given by (5.5.17).
2. Compute m^* from

$$\frac{\partial}{\partial m}\hat{v}(\tilde{\Delta}^*,m^*) = 0,$$

where $\hat{v}(\Delta,m)$ is the approximation of $v(\Delta,m)$, that results from the approximations given in chapter 1 and approximation 5.5.1.

Alternatively we can minimize $\hat{v}(\Delta,m)$ as a function of Δ and m subject to $\Delta,m\geq 0$. This procedure is more computer-time-consuming, but yields more accurate results in the case of a small shortage cost rate p.

In the next section we also give numerical results concerning problem Pb 2. The numerical results obtained by our method will be compared with those obtained by a diffusion approximation studied in Vickson [1982].

5.6. Numerical results and conclusions.

In the introduction to this chapter we posed three questions: Is it possible to obtain accurate approximations for the holding and switching costs? Is the approximate (m,M)-rule with M-m predetermined by the EOQ-formula (1.5.1) a good rule with respect to average costs? Is the optimal M-m insensitive to the required service level? In section 5.5 we

proved that the last question can be answered in the affirmative provided the switching level m is sufficiently large. This answer to the last question has an important consequence for the second question. Since it is easy to verify that the solution $\tilde{\Delta}^*$ to equation (5.5.17) is not equal to the difference Δ_{EOQ} resulting from (1.5.1), one may wonder whether the second question can have a positive answer.

Throughout this section we restrict to the backlog model, since the approximations 5.5.1-5.5.4 hold for all models discussed in the chapters 1 to 4. Also, in view of the fact that $\tilde{\Delta}^*$ is the same for each of the service measures considered (see section 5.5), we only consider the β-service measure requiring that the fraction of demand being met directly from stock on hand equals β. After having discussed the answer to the above three questions we compare at the end of this section our approximations with a diffusion approximation.

Let us address the first question. In table 5.6.1 we give the approximate and the actual values of the average on-hand inventory for the (m,M)-rules that are given in table 1.5.1. In all these examples $\lambda=1$ and $E[D]=1$, while M-m is determined by formula (1.5.1) with K=25 and h=1. We assume deterministic demand ($c_D^2=0$) and gamma demand with $c_D^2=1/3$, 2/3 and 2. The numerical results allow the conclusion that the approximation to the average on-hand inventory shows an excellent performance.

In table 5.6.2 we show the convergence of Δ^* to $\tilde{\Delta}^*$ as the service level β approaches 1. Here Δ^* is the approximate optimal difference M-m that is computed by using approximation 5.5.2 for the expected total costs per cycle. In all examples we have chosen $\lambda=1$, $E[D]=1$, h=1 and K=25.
Again we used deterministic and gamma distributions to represent the distribution of the demand D. For deterministic demand and $\pi_1=0$ approximation 5.5.4 is not valid, since the undershoot is deterministic and depending on Δ. Hence the insensitivity result does not hold for this particular case. Still we computed $\tilde{\Delta}^*$ from $z(\Delta)=0$. As a rule of thumb we state that $\Delta^* \cong \tilde{\Delta}^*$ when the required service level is at least 99%. Of course this does not hold when production rate π_2 is extremely high, causing m to be small.

Table 5.6.1. The approximate average inventories and their actual values.

			$c_D^2=0$			$c_D^2=1/3$		
π_1	π_2	β	m	V_{app}	V_{act}	m	V_{app}	V_{act}
-0.5	1.25	0.95	5.57	5.45	5.46 (4)	8.09	7.37	7.42 (5)
-0.5	2	0.95	1.02	3.43	3.43 (1)	1.82	4.12	4.12 (1)
-0.5	5	0.95	0.26	4.00	4.00 (1)	0.57	4.31	4.31 (1)
0	1.25	0.95	5.74	5.49	5.50 (5)	8.05	7.39	7.37 (6)
0	2	0.95	1.44	3.46	3.46 (1)	1.87	4.04	4.05 (2)
0	5	0.95	0.45	3.65	3.65 (1)	0.65	3.97	3.97 (1)
0.5	1.25	0.95	5.15	5.49	5.48 (4)	7.52	7.43	7.44 (8)
0.5	2	0.95	0.76	3.13	3.12 (2)	1.51	3.93	3.93 (2)
0.5	5	0.95	0.04	3.08	3.09 (2)	0.33	3.54	3.54 (2)
-0.5	1.25	0.99	9.31	9.11	9.13 (5)	13.26	12.44	12.49 (8)
-0.5	2	0.99	2.30	4.70	4.71 (2)	3.74	6.01	6.02 (2)
-0.5	5	0.99	0.86	4.59	4.59 (1)	1.65	5.38	5.38 (1)
0	1.25	0.99	9.47	9.15	9.19 (5)	13.22	12.45	12.50 (8)
0	2	0.99	2.73	4.73	4.74 (1)	3.79	5.94	5.93 (2)
0	5	0.99	0.96	4.16	4.16 (2)	1.72	5.03	5.04 (1)
0.5	1.25	0.99	8.88	9.15	9.18 (5)	12.69	12.50	12.51 (8)
0.5	2	0.99	2.05	4.40	4.40 (2)	3.43	5.82	5.81 (2)
0.5	5	0.99	0.66	3.70	3.70 (2)	1.41	4.61	4.62 (2)
			$c_D^2=2/3$			$c_D^2=2$		
π_1	π_2	β	m	V_{app}	V_{act}	m	V_{app}	V_{act}
-0.5	1.25	0.95	10.65	9.32	9.39(11)	21.02	17.19	17.21(22)
-0.5	2	0.95	2.72	4.89	4.89 (2)	6.71	8.35	8.33 (4)
-0.5	5	0.95	0.96	4.70	4.70 (1)	3.01	6.71	6.71 (2)
0	1.25	0.95	10.58	9.33	9.29(12)	20.85	17.20	17.47(22)
0	2	0.95	2.77	4.84	4.84 (2)	6.71	8.33	8.34 (4)
0	5	0.95	1.06	4.40	4.40 (1)	3.16	6.54	6.54 (2)
0.5	1.25	0.95	9.92	9.39	9.37(12)	19.68	17.80	17.32(25)
0.5	2	0.95	2.33	4.78	4.78 (3)	5.88	8.44	8.46 (5)
0.5	5	0.95	0.69	4.07	4.06 (3)	2.53	6.50	6.49 (4)
-0.5	1.25	0.99	17.26	15.80	15.79(18)	33.39	29.32	29.45(25)
-0.5	2	0.99	5.28	7.43	7.43 (2)	11.91	13.49	13.46 (4)
-0.5	5	0.99	2.51	6.24	6.24 (1)	6.43	10.11	10.12 (2)
0	1.25	0.99	17.19	15.81	15.80(12)	33.23	29.32	29.31(35)
0	2	0.99	5.33	7.37	7.36 (2)	11.92	13.47	13.46 (5)
0	5	0.99	2.61	5.94	5.94 (2)	6.58	9.95	9.95 (2)
0.5	1.25	0.99	16.53	15.87	15.94(13)	32.05	29.43	29.37(25)
0.5	2	0.99	4.89	7.32	7.30 (3)	11.08	13.58	13.58 (5)
0.5	5	0.99	2.23	5.60	5.61 (3)	5.95	9.90	9.87 (4)

Table 5.6.2. Convergence of Δ^* to $\tilde{\Delta}^*$.

		$c_D^2=0$				$c_D^2=1/3$			
π_1	π_2	$\beta=.90$	$\beta=.95$	$\beta=.99$	$\tilde{\Delta}^*$	$\beta=.90$	$\beta=.95$	$\beta=.99$	$\tilde{\Delta}^*$
-0.5	1.25	5.116	5.116	5.116	5.116	5.566	5.566	5.566	5.566
-0.5	2	6.007	5.990	5.985	5.985	6.313	6.300	6.300	6.299
0	1.25	4.558	5.000	4.930	5.000	5.202	5.202	5.202	5.202
0	2	5.015	5.000	5.000	5.000	5.608	5.599	5.599	5.599
0.5	1.25	4.620	4.620	4.620	4.620	4.926	4.926	4.926	4.926
0.5	2	4.606	4.544	4.564	4.564	4.771	4.757	4.757	4.756
		$c_D^2=2/3$				$c_D^2=2$			
π_1	π_2	$\beta=.90$	$\beta=.95$	$\beta=.99$	$\tilde{\Delta}^*$	$\beta=.90$	$\beta=.95$	$\beta=.99$	$\tilde{\Delta}^*$
-0.5	1.25	5.936	5.936	5.936	5.936	7.004	7.004	7.004	7.004
-0.5	2	6.614	6.606	6.605	6.605	7.600	7.625	7.638	7.638
0	1.25	5.505	5.505	5.505	5.505	6.323	6.323	6.323	6.323
0	2	5.845	5.838	5.837	5.837	6.554	6.579	6.593	6.593
0.5	1.25	5.163	5.163	5.163	5.163	5.766	5.765	5.765	5.765
0.5	2	4.938	4.930	4.928	4.928	5.393	5.420	5.438	5.438

Table 5.6.2 shows that the optimal difference Δ^* is indeed insensitive to the required service level when this level is sufficiently high. Denote by g_{EOQ} and $\tilde{g}^*$ the average switching and holding costs of the (m,M)-rules that are obtained for the β-service level requirement when using (1.5.1) for M-m respectively using $\tilde{\Delta}^*$ for M-m. In table 5.6.3 we compare the average costs g_{EOQ} and $\tilde{g}^*$ with the minimal average switching and holding costs g^* associated with the optimal (m,M)-rule obtained by solving problem Pb 1. The costs g_{EOQ}, $\tilde{g}^*$ and g^* are computed from the approximations 5.5.2 and 5.5.3. In all examples we have chosen $\lambda=1$, $E[D]=1$, $h=1$ and $K=25$.

From table 5.6.2 we found that $\tilde{\Delta}^*$ and Δ^* differ only slightly even for $\beta=0.9$ and so it is no surprise that $\tilde{g}^*$ and g^* are almost identical. However, Δ_{EOQ} and $\tilde{\Delta}^*$ may differ significantly, cf. tables 1.5.1 and 5.6.2.

Nevertheless the results in table 5.6.3 show that the $(m,m+\Delta_{EOQ})$-rule performs quite well. We found that the relative error is usually below 5%,

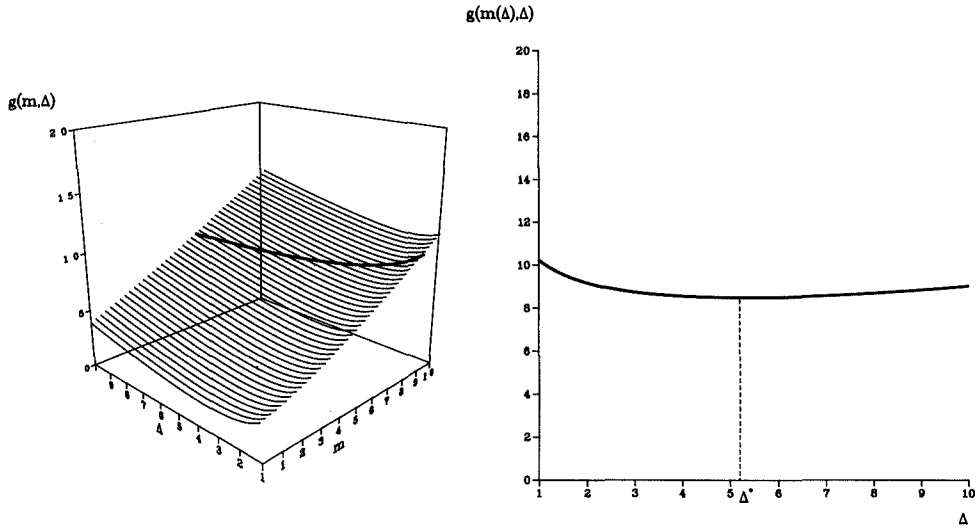
Table 5.6.3. Comparison of g_{EOQ}, $\tilde{g}^*$ and g^*.

			$c_D^2=0$			$c_D^2=1/3$		
π_1	π_2	β	g_{EOQ}	$\tilde{g}^*$	g^*	g_{EOQ}	$\tilde{g}^*$	g^*
-0.5	1.25	0.90	5.4206	5.2011	5.2011	6.7177	6.4478	6.4478
-0.5	2	0.90	5.4938	5.4738	5.4738	5.8589	5.8120	6.8119
-0.5	1.25	0.95	6.9366	6.7171	6.7171	8.8155	8.5456	8.5456
-0.5	2	0.95	6.0127	5.9915	5.9915	6.6553	6.6094	6.6094
0	1.25	0.90	5.2276	5.1298	5.1298	6.5930	6.3875	6.3875
0	2	0.90	5.0223	5.0223	5.0223	5.4540	5.4285	5.4285
0	1.25	0.95	6.7436	6.6458	6.6458	8.6908	8.4853	8.4853
0	2	0.95	5.5455	5.5455	5.5455	6.2500	6.2250	6.2250
0.5	1.25	0.90	5.2483	5.0488	5.0488	6.5312	6.3184	6.3184
0.5	2	0.90	4.4793	4.4556	4.4555	4.9175	4.8805	4.8805
0.5	1.25	0.95	6.7643	6.5648	6.5648	8.6290	8.4162	8.4162
0.5	2	0.95	4.9996	4.9793	4.9793	5.7136	5.6776	5.6776
			$c_D^2=2/3$			$c_D^2=2$		
π_1	π_2	β	g_{EOQ}	$\tilde{g}^*$	g^*	g_{EOQ}	$\tilde{g}^*$	g^*
-0.5	1.25	0.90	8.0390	7.7367	7.7367	13.4232	13.0685	13.0685
-0.5	2	0.90	6.3145	6.2359	6.2359	8.5109	8.3205	8.3205
-0.5	1.25	0.95	10.7203	10.4180	10.4180	18.4446	18.0899	18.0899
-0.5	2	0.95	7.3802	7.3024	7.3024	10.6643	10.4706	10.4706
0	1.25	0.90	7.9016	7.6788	7.6788	13.2474	13.0133	13.0133
0	2	0.90	5.9154	5.8716	5.8716	8.1011	8.0022	8.0022
0	1.25	0.95	10.5829	10.3601	10.3601	18.2688	18.0347	18.0347
0	2	0.95	6.9811	6.9379	6.9379	10.2536	10.1521	10.1521
0.5	1.25	0.90	7.8406	7.6257	7.6257	13.2032	13.0118	13.0118
0.5	2	0.90	5.4275	5.3779	5.3779	7.7527	7.6797	7.6796
0.5	1.25	0.95	10.5219	10.3071	10.3071	18.2247	18.0334	18.0333
0.5	2	0.95	6.4934	6.4446	6.4446	9.9031	9.8275	9.8275

cf. also De Kok et al [1984] and De Kok [1985]. This indicates that the cost function $g(m(\Delta),\Delta)$ with $m(\Delta)$ uniquely determined by the service level constraint, is extremely flat arounds its minimum. Figure 5.6.1 shows a typical picture of $g(m,\Delta)$ as a function of both m and Δ and of $g(m(\Delta),\Delta)$ as a function of Δ. It is interesting to note that the difference between Δ_{EOQ}

and Δ^* is largest when π_2 gets close to $\lambda E[D]$; the same holds for g_{EOQ} and g^*. In contrast with this observation $\tilde{g}^*$ gets closer to g^* as π_2 decreases to $\lambda E[D]$. The following explanation of this finding can be given. Assuming that the service level β is fixed, we first note that the switching level m must increase as the production rate π_2 approaches the mean demand per unit time. Then the approximation 5.5.4 for the average costs leading to $\tilde{\Delta}^*$ is nearly the same as the approximation for the average costs that follows from approximations 5.5.2 and 5.5.3 leading to Δ^*. Hence in that case $\tilde{\Delta}^*$ must be close to Δ^* and so $\tilde{g}^*$ to g^*.

FIGURE 5.6.1.



Concluding, the (m,M)-rule using Δ_{EOQ} shows a good performance in costs in most cases. Nevertheless we recommend the use of $\tilde{\Delta}^*$ because the resulting (m,M)-rules are approximately optimal for the whole range of parameter values. The computation of $\tilde{\Delta}^*$ is not more difficult than the computation of the number δ that has to be computed anyway.

As we advocate the use of $\tilde{\Delta}^*$ instead of Δ_{EOQ}, it is interesting to plot $\tilde{\Delta}^*$ as a function of the model parameters π_1, π_2 and c_D^2. In table 5.6.4 we give the values of $\tilde{\Delta}^*$ for a number of combinations of these model parameters. As before we have chosen $\lambda=1$, $E[D]=1$, $h=1$ and $K=25$. In all cases we fitted a gamma distribution to the first two moments of the demand size distribution.

Table 5.6.4. The number $\tilde{\Delta}^*$ as a function of π_1, π_2 and c_D^2.

π_1	π_2 \ c_D^2	0.5		2	4	8	16
0	1.25	5.36	5.76	6.32	7.11	8.43	11.51
0	2	5.72	6.05	6.59	7.46	9.16	13.35
0	5	6.38	6.61	7.10	8.17	10.73	17.21
0.5	1.25	5.05	5.35	5.77	6.31	7.27	9.74
0.5	2	4.85	5.08	5.44	6.04	7.40	11.02
0.5	5	4.84	5.00	5.35	6.20	8.51	14.44

We summarize the following properties of $\tilde{\Delta}^*$. The value of $\tilde{\Delta}^*$ increases as c_D^2 increases, as is in fact required by condition 1.3.1. This finding provides an additional argument to use $\tilde{\Delta}^*$ rather than Δ_{EOQ} (we found that usually $\Delta_{EOQ} \leq \tilde{\Delta}^*$). The value of $\tilde{\Delta}^*$ is, like Δ_{EOQ}, monotonically decreasing in π_1. As opposed to this finding we have that Δ_{EOQ} is monotonically increasing in π_2, whereas $\tilde{\Delta}^*$ is not monotonic in π_2. It is intuitively clear that $\tilde{\Delta}^*$ should be increasing in K and decreasing in h. This is indeed true as can be verified from (5.5.17).

Finally in tables 5.6.5 and 5.6.6 we compare our approximate results for the compound Poisson process with the results for a diffusion approximation. We note that the diffusion process approximation is completely determined by the first two moments of the demand per unit time. Assuming that the inventory is controlled by an (m,M)-rule and the production is governed by one out of two production rates π_1 and π_2, it follows that the diffusion demand process induces a two-mode diffusion inventory process. Using results from Vickson [1982] for the backlog model it is easy to find expressions for the average length of a regeneration cycle, the average on-hand inventory and the average backlog at an arbitrary point in time.

Now we address the question whether a diffusion process approximation

to the compound Poisson demand process yields acceptable results for the average on-hand inventory. Therefore we consider a number of examples with $\lambda=1$, $E[D]=1$, $h=1$ and $K=25$. It follows from the results in table 5.6.5 that the diffusion process approximation yields an overestimate of the value of the average on-hand inventory. This is probably caused by the fact that the undershoot of the switching level m is zero for the diffusion process approximation.

Table 5.6.5. Average inventories for compound Poisson and diffusion demand.

		$c_D^2=0.5$				$c_D^2=2$			
π_1	π_2	m	M	V_{Pois}	V_{diff}	m	M	V_{Pois}	V_{diff}
0	1.25	9.31	12.47	8.36	8.71	20.85	24.01	17.20	18.05
0	2	2.31	7.31	4.43	4.82	6.71	11.71	8.33	9.21
0	5	0.85	7.17	4.18	4.57	3.16	9.49	6.54	7.45
0.5	1.25	8.72	11.61	8.41	8.73	19.68	22.56	17.30	18.24
0.5	2	1.91	6.00	4.35	4.71	5.88	9.96	8.44	9.42
0.5	5	0.50	5.22	3.80	4.17	2.53	7.25	6.50	7.51
0	1.25	15.20	18.36	14.13	14.54	33.23	36.39	29.32	30.32
0	2	4.55	9.55	6.64	7.05	11.92	16.92	13.47	14.42
0	5	2.16	8.48	5.48	5.88	6.58	12.91	9.95	10.87
0.5	1.25	14.61	17.50	14.18	14.56	32.05	34.94	29.43	30.51
0.5	2	4.15	8.23	6.56	6.94	11.08	15.17	13.58	14.63
0.5	5	1.81	6.53	5.10	5.48	5.95	10.67	9.90	10.94

Next we assume that a fixed cost $K\geq 0$ is incurred each time the production rate is switched from π_1 to π_2. Also, a holding cost at rate $h.x$ is incurred when the stock on hand equals $x\geq 0$ and a penalty cost at rate $p.z$ is incurred when the backlog equals $z\geq 0$. For this cost structure and a diffusion demand process Vickson [1982] shows that an (m,M)-rule is average-cost optimal among all reasonable policies. Also, he gives a numerical procedure to determine the optimal (m,M)-policy when considering the switching, holding and penalty costs only. This numerical procedure may be simplified considerably. By deriving directly an explicit expression for the average cost of a given (m,M)-rule, it is easy to find the optimal values of m and M. Moreover, it follows that the optimal value of $\Delta=M-m$ is independent of the penalty cost rate p. Using this result, the optimal $(m,m+\Delta)$-rule can be computed by using the two-step-procedure described below problem Pb 2. We note that the optimal value of m may be negative. The approximations given in this monograph for the compound

Poisson demand case are valid only when $m \geq 0$. The results from table 5.6.6 reveal that a diffusion process approximation to the compound Poisson demand process may lead to policies being quite different to those obtained by our renewal-theoretic approach, where the diffusion process approximation for the minimal average costs may underestimate rather dramatically the actual minimum costs. For the compound Poisson demand case the optimal $(m,m+\Delta)$-rule is determined by the algorithm described below Pb 1.

Table 5.6.6. Optimal $(m,m+\Delta)$-rule for compound Poisson and diffusion demand.

$c_D^2=0.5$		compound Poisson			diffusion		
π_2	p	m	Δ	$g(m,\Delta)$	m	Δ	$g(m,\Delta)$
1.25	12	6.04	5.36	10.54	4.59	5.76	9.09
2	12	0.58	5.73	6.36	-0.13	5.80	5.67
1.25	16	7.02	5.36	11.52	5.40	5.76	9.90
2	16	0.96	5.72	6.74	0.07	5.80	5.88
$c_D^2=2$		compound Poisson			diffusion		
π_2	p	m	Δ	$g(m,\Delta)$	m	Δ	$g(m,\Delta)$
1.25	12	14.33	6.32	20.39	10.82	7.17	16.60
2	12	2.56	6.56	9.69	0.56	6.62	7.26
1.25	16	16.39	6.32	22.45	12.43	7.17	18.21
2	16	3.41	6.57	10.55	0.97	6.62	7.66

6. A PRODUCTION-INVENTORY MODEL WITH POSITIVE SETUP TIME.

So far we have assumed that it takes no time to switch from one production rate to another. In some practical applications it may be more appropriate to assume that a positive switch time is needed to adjust the production rate, especially when switching from a low to a high production rate.

In this chapter we restrict to the case that the low production rate is equal to zero. In other words, the production facility is either on or off. As before the inventory is controlled by an (m,M)-rule with $0 \leq m \leq M$ with the only difference that as soon as the inventory level drops below m then the production facility is reactivated and production continues after a positive (possibly stochastic) setup time.

We consider both the backlog and the lost-sales model. Again we focus on the derivation of approximations for service levels with respect to some given service measure. We sequentially determine M-m and m such that a prespecified service level is achieved.

The chapter is organized as follows. In section 6.1 we describe the model in detail and derive expressions for the service measures in terms of a number of basic functions. Most of these basic functions are already known from chapters 1 and 2. In section 6.2 we derive two-moment approximations based on practical distributions. The behaviour of the inventory process during the setup time is studied in section 6.3. Combining the results obtained in the sections 6.1-6.3, we obtain the desired approximations for the service measures. In section 6.4 we derive approximations for holding and switching costs under some cost structure. Section 6.5 concludes this chapter with the presentation and discussion of numerical results.

6.1. Model and service measures.

In this section we first give a detailed description of the models to be considered. We assume that customers arrive according to a Poisson process with rate λ. The demands of the customers are independent random variables with common distribution function F with $F(0)=0$. The demands are independent of the arrival process itself. We deal with both the model in which excess demand is backlogged and the model in which excess demand is lost.

At any point in time the production is either on or off. If production

is on then items are continually added to inventory at rate π_2. Letting D denote the demand of a single customer we assume that

(6.1.1) $\qquad \pi_2 > \lambda E[D].$

The inventory is controlled by an (m,M)-rule. If the inventory level becomes as high as M, the production is immediately stopped and the production facility is shut down. The production facility is reactivated again as soon as the inventory level drops below m. After a setup time T the production continues at rate π_2. We assume that the random variable T is independent of the demand process and the current inventory.

We recapitulate a number of definitions given in the chapters 1 and 2. Define for any $t \geq 0$,

$N(t)$:= the number of customers that arrive in $(0,t]$.

$V(t)$:= the total demand in $(0,t]$.

$X(t)$:= the inventory level at time t.

$B(t)$:= the amount of demand in $(0,t]$ that cannot be met directly from stock on hand.

$Q(t)$:= the number of stockout occurrences in $(0,t]$.

$S(t)$:= the number of customers arriving in $(0,t]$ whose demands cannot be met directly from stock on hand.

$C(t) := -\int_0^t X(s) 1_{\{X(s)<0\}} ds$ = the cumulative backlog at time t.

We say that a stockout occurs if the inventory level drops from a positive value to a non-positive value. Then we consider the following service measures.

(i) <u>α-service measure</u>.

the long-run average number of stockout occurrences per unit time,

$$\lim_{t\to\infty} \frac{Q(t)}{t} .$$

(ii) β-service measure.
the long-run fraction of demand that cannot be met directly from stock on hand,

$$\lim_{t\to\infty} \frac{B(t)}{V(t)} .$$

(iii) γ-service measure.
the long-run fraction of customers whose demand cannot be met directly from stock on hand,

$$\lim_{t\to\infty} \frac{S(t)}{N(t)} .$$

(iv) δ-service measure.
the long-run average backlog at an arbitrary point in time,

$$\lim_{t\to\infty} \frac{C(t)}{t} .$$

Since the inventory is controlled by an (m,M)-rule the inventory process is regenerative. We take as regeneration epochs the epochs at which the level M is reached from below. Unless stated otherwise we assume that epoch 0 is a regeneration epoch. Next we define the following random variables,

T := the next epoch at which the level M is reached from below.

Also, we associate with the cycle (0,T] the random variables

$$N := N(T), \quad V := V(T), \quad B := B(T), \quad Q := Q(T), \quad S := S(T),$$
$$C := C(T).$$

It follows from the theory of regenerative processes that the following equalities hold with probability 1,

$$\lim_{t\to\infty} \frac{Q(t)}{t} = \frac{E[Q]}{E[T]}, \quad \lim_{t\to\infty} \frac{B(t)}{V(t)} = \frac{E[B]}{E[V]}$$

(6.1.2)

$$\lim_{t\to\infty} \frac{S(t)}{N(t)} = \frac{E[S]}{E[N]}, \quad \lim_{t\to\infty} \frac{C(t)}{t} = \frac{E[C]}{E[T]}.$$

Due to the compound Poisson demand process we have that

(6.1.3) $\quad E[N] = \lambda E[T], \quad E[V] = \lambda E[D]E[T].$

Hence it suffices to find expressions for E[T], E[Q], E[B], E[S] and E[C]. Towards this end we introduce a number of basic functions. We distinguish between three types of basic functions. Those associated with the production facility being shut down and not yet reactivated, those associated with the inventory process during the setup time and those associated with the production being on.

First we define the basic functions associated with the production facility being shut down and not yet reactivated. Under the condition that at epoch 0 the inventory level equals x+m, $x \geq 0$, and the production is off, let

$t_1(x)$:= the expected time until the inventory level drops below m and the production facility is reactivated.

$p(x,u) = P\{U(x)>u\}, \quad u \geq 0,$

where the random variable U(x) is defined by

U(x) := the undershoot (included any shortage) of the inventory level m at the demand epoch at which the inventory level drops below m and the production facility is reactivated.

Expressions for $t_1(x)$, p(x,u) and E[U(x)] are given by (1.3.33)-(1.3.35) for the case of x=0 and by approximations 1.3.1-1.3.3 when x satisfies condition 1.3.1.

Next we define the basic functions associated with the inventory process during the setup time. Let a "dissatisfied" customer be a customer whose demand cannot be met directly from stock on hand. Then we define

$s(T)$:= the expected number of dissatisfied customers arriving during the setup time (including the customer initiating the setup time if this customer is a dissatisfied one).

$c(T)$:= the expected amount by which the cumulative backlog is augmented during the setup time T.

For ease of notation we suppressed in the above defined basic functions the dependency on m and M. Expressions for these basic functions will be derived in section 6.3.

Finally we define the basic functions associated with the production being on. Assuming that at epoch 0 the inventory level equals $x \leq M$ and the production rate π_2 is used, we define

$t_2(x)$:= the expected time until the inventory reaches the level M.

$b(x)$:= the expected amount of demand that is not met directly from stock on hand until the inventory reaches the level M (excluding any shortage existing at epoch 0).

$q(x)$:= the probability that the inventory level becomes non-positive before the inventory reaches the level M.

$c(x)$:= the expected cumulative backlog at the time at which the inventory reaches the level M.

Expressions for these basic functions are given in the chapters 1 and 2 by the approximations 1.4.1-1.4.3 for the backlog model and the approximations 2.3.1-2.3.3 for the lost-sales model. It follows from equations (1.4.1), (2.3.1), (2.3.22) and (2.3.23) that

$$t_{2,L}(x) = t_{2,B}(x) - \frac{b_B(x)}{\pi_2}, \qquad 0 \leq x \leq M. \tag{6.1.4}$$

$$b_L(x) = \frac{\pi_2 - \lambda E[D]}{\pi_2} b_B(x), \qquad 0 \leq x \leq M. \tag{6.1.5}$$

$$q_L(x) = q_B(x), \qquad 0 \leq x \leq M. \tag{6.1.6}$$

Here the subscripts B and L refer to the backlog and lost-sales model respectively.

Now we define the following random variables,

ξ := the total demand during the setup time T

$\eta := \xi + U(M-m)$.

Hence η is the sum of the total demand during the setup time and the undershoot of level m (including any shortage). We also note that ξ is independent of m and M. We further define

$$F_\xi(x) = P\{\xi \leq x\}, \quad x \geq 0$$

$$F_\eta(x) = P\{\eta \leq x\}, \quad x \geq 0.$$

Again for ease of notation we have suppressed the dependency of η on M-m.

Proceeding along the same lines as in the derivation of relations (1.2.8)-(1.2.11) and (2.1.4)-(2.1.6) we obtain

$$E[T_B] = t_1(M-m) + E[T] + \int_0^\infty t_{2,B}(m-y)dF_\eta(y) \tag{6.1.7}$$

$$E[B_B] = \int_0^\infty b_B(m-y)dF_\eta(y) + \int_0^\infty (y-m)dF_\eta(y) \tag{6.1.8}$$

$$E[Q_B] = (1-q_B(0))^{-1}\{1-F_\eta(m) + \int_0^m q_B(m-y)dF_\eta(y)\} \tag{6.1.9}$$

$$E[S_B] = s(T) + \frac{\lambda}{\pi_2} E[B_B] + E[Q_B] - (1-F_\eta(m)) \tag{6.1.10}$$

$$E[C_B] = c(T) + \int_0^\infty c_B(m-y)dF_\eta(y) \tag{6.1.11}$$

$$E[T_L] = t_1(M-m) + E[T] + \int_0^\infty t_{2,L}(m-y)dF_\eta(y) \tag{6.1.12}$$

$$E[B_L] = \int_0^m b_L(m-y)dF_\eta(y) + b_L(0)(1-F_\eta(m)) + \int_m^\infty (y-m)dF_\eta(y) \tag{6.1.13}$$

$$E[Q_L] = (1-q_L(0))^{-1}\{1-F_\eta(m) + \int_0^m q_L(m-y)dF_\eta(y)\} \tag{6.1.14}$$

$$E[S_L] = s(T) + E[Q_L] - (1-F_\eta(m)). \tag{6.1.15}$$

The expressions for $E[S_B]$ and $E[S_L]$ need some more explanation. Let us consider the backlog model. The number of dissatisfied customers in a cycle equals the sum of the number of dissatisfied customers arriving in the setup

time and the number of dissatisfied customers arriving while the production is on. The expected number of dissatisfied customers arriving in the setup time is by definition equal to $s(T)$. The expected number of dissatisfied customers arriving while production is on equals the expected number of stockout occurrences while production is on plus the expected number of customers arriving while the inventory is negative and production is on. Since any shortage is produced at rate π_2 it follows that the expected time that the inventory is negative while production is on equals $E[B_B]/\pi_2$. Using "Poisson arrivals see time averages" it follows that the expected number of customers arriving while the inventory is negative and production is on equals $\lambda E[B_B]/\pi_2$. The expected number of stockout occurrences while production is on equals the number of stockout occurrences during a cycle minus the expected number of stockout occurrences during the setup time. The latter number equals $(1-F_\eta(m))$ since during the setup time T the inventory decreases with jumps and if $\eta \geq m$ then the inventory has crossed the level 0 exactly once. A combination of the above arguments yields (6.1.10). Analogously we derived (6.1.15).

Next we express $E[T_L]$, $E[B_L]$, $E[Q_L]$ and $E[S_L]$ in terms of $E[T_B]$, $E[B_B]$, $E[Q_B]$ and $E[S_B]$. Equation (1.4.19) states that

$$(6.1.16) \qquad b_B(x) = \frac{-\lambda E[D]x}{\pi_2 - \lambda E[D]} + b_B(0), \quad x \leq 0.$$

Then equations (6.1.4)-(6.1.10) and equations (6.1.12)-(6.1.16) together imply the following exact relations

$$(6.1.17) \qquad E[T_L] = E[T_B] - E[B_B]/\pi_2$$

$$(6.1.18) \qquad E[B_L] = \frac{\pi_2 - \lambda E[D]}{\pi_2} E[B_B]$$

$$(6.1.19) \qquad E[Q_L] = E[Q_B]$$

$$(6.1.20) \qquad E[S_L] = E[S_B] - \frac{\lambda}{\pi_2} E[B_B].$$

It is important to note that, unlike in the case of no setup time, $E[Q_L] \neq E[S_L]$, since in the lost-sales model at most one stockout can occur during the setup time. If a stockout has occurred during the setup time all customers arriving after this stockout occurrence, but before production is on again, are dissatisfied. Their demand is totally lost. However, their

arrivals do not cause a stockout.

Due to relations (6.1.17)-(6.1.20) we may restrict further attention to the backlog model. Still we need (approximate) expressions for $F_\eta(x)$, $s(T)$ and $c(T)$. Also, an approximation for $F_\xi(x)$ must be found. Section 6.2 deals with approximations for $F_\eta(x)$ and $F_\xi(x)$, whereas $s(T)$ and $c(T)$ are approximated in section 6.3.

6.2. Approximations for $F_\xi(x)$ and $F_\eta(x)$.

In this section we propose approximations for the distribution function of the demand in the setup time ξ and the distribution function of η, being the sum of the undershoot $U(M-m)$ of m (including any shortage) and the demand in the setup time. These approximations are based on the first two moments of ξ and η. The reason for this rather crude approach is that in general it is impossible to obtain tractable exact expressions for $F_\xi(x)$ and $F_\eta(x)$. Fortunately, numerical investigations show that the approximations based on only the first two moments perform satisfactorily.

It is well-known (cf. Ross [1970]) that the first two moments of ξ are given by

$$E[\xi] = \lambda E[D]E[T] \tag{6.2.1}$$

$$E[\xi^2] = \lambda E[D^2]E[T] + (\lambda E[D])^2 E[T^2]. \tag{6.2.2}$$

We also define the squared coefficient of variation of ξ,

$$c_\xi^2 := \frac{(E[\xi^2]-(E[\xi])^2)}{(E[\xi])^2} .$$

The value of c_ξ^2 determines whether the demand in the setup time is non-erratic ($c_\xi^2 \leq 1$) or erratic ($c_\xi^2 > 1$). In approximating $F_\xi(x)$ we also distinguish between these two cases.

We approximate $F_\xi(x)$ by $\hat{F}_\xi(x)$, which is defined by

$$\hat{F}_\xi(x) := p_1\Big(1- \sum_{j=0}^{r} e^{-\beta_1 x} \frac{(\beta_1 x)^j}{j!}\Big) + (1-p_1)\Big(1- \sum_{j=0}^{n} e^{-\beta_2 x} \frac{(\beta_2 x)^j}{j!}\Big),$$

$$x \geq 0.$$

In the determination of the constants n, r, p_1, β_1 and β_2, we distinguish between the cases $c_\xi^2 \leq 1$ and $c_\xi^2 > 1$.

Case (i): $c_\xi^2 \leq 1$.
The integers n and r are uniquely determined by

$$\frac{1}{n} < c_\xi^2 \leq \frac{1}{n-1}, \qquad r = n-1,$$

while

$$p_1 := \frac{1}{1+c_\xi^2}[nc_\xi^2 - \{n(1+c_\xi^2) - n^2c_\xi^2\}^{\frac{1}{2}}]$$

$$\beta_1 := \frac{n-p_1}{E[\xi]}, \qquad \beta_2 := \beta_1.$$

Case (ii): $c_\xi^2 > 1$.
The constants n, r, p_1, β_1 and β_2 are now defined by

$$n := 1, \qquad r := 1$$

$$\beta_1 := \frac{2}{E[\xi]}\left[1 + \left\{\frac{c_\xi^2 - \frac{1}{2}}{c_\xi^2+1}\right\}^{\frac{1}{2}}\right], \qquad \beta_2 := \frac{4}{E[\xi]} - \beta_1$$

$$p_1 := \frac{\beta_1(\beta_2 E[\xi]-1)}{\beta_2-\beta_1}.$$

In the case of $c_\xi^2 \leq 1$ we approximate $F_\xi(x)$ by a mixture of an Erlang-n and an Erlang-(n-1) distribution function with the same scale parameter. In the case of $c_\xi^2 > 1$ we fit a hyperexponential distribution function with the same first three moments as the gamma distribution. It is easily verified that the constants n, r, p_1, β_1 and β_2 are such that $\hat{F}_\xi(x)$ has the same first two moments as $F_\xi(x)$.

The approximation $\hat{F}_\eta(x)$ of $F_\eta(x)$ is obtained in exactly the same way. Since $\eta = U(M-m)+\xi$ we obtain

$$E[\eta] = E[U(M-m)] + E[\xi] \tag{6.2.3}$$

$$E[\eta^2] = E[U^2(M-m)] + 2E[\xi]E[U(M-m)] + E[\xi^2]. \tag{6.2.4}$$

Letting c_η^2 denote the squared coefficient of variation of η,

$$c_\eta^2 := \frac{E[\eta^2]-(E[\eta])^2}{(E[\eta])^2},$$

we replace $E[\xi]$ and c_ξ^2 by $E[\eta]$ and c_η^2, respectively, in the above given formulae for n, r, p_1, β_1 and β_2 to obtain $\hat{F}_\eta(x)$, which also is a two-moment approximation of $F_\eta(x)$.

Due to the form of the distribution function $\hat{F}_\eta(x)$ and the approximations given for $t_{2,B}(x)$, $b_B(x)$, $q_B(x)$ and $c_B(x)$ in chapter 1, it is now an easy matter to obtain expressions for the integrals on the right-hand side of (6.1.7)-(6.1.11). It remains to find expressions for $s(T)$ and $c(T)$. In the computation of these expressions we apply the approximation $\hat{F}_\xi(x)$ of $F_\xi(x)$.

6.3. The basic functions associated with setup time T.

As said before the basic functions $s(T)$ and $c(T)$ depend on both m and M. More precisely, they depend on m and M-m. The dependence on M-m is induced by the undershoot distribution p(M-m,u). We assumed that T is independent of the compound Poisson demand process. Therefore our approach will be as follows. We first derive expressions for $s(T)$ and $c(T)$ conditional on the event $\{T=t\}$ for some $t\geq 0$. Using these expressions we derive the unconditional expressions for $s(T)$ and $c(T)$.

Let us assume that $T=t$ for some $t\geq 0$. Let us further assume that production is <u>always</u> off and at epoch 0 the inventory equals x. Define

$s(x,t)$:= the expected number of customers arriving in (0,t] whose demands cannot be met directly from stock on hand.

$c(x,t)$:= the expected cumulative backlog at epoch t.

First we derive an expression for s(x,t). It follows from "Poisson arrivals see time averages" that

(6.3.1) $\qquad s(x,t) = \lambda t, \quad x \leq 0.$

So let us assume x>0. Then we define the following random variables.

T(x) := the time until the inventory drops below 0.

$\bar{T}(x,t)$:= the time that the inventory is negative during $(0,t]$.

$\xi(t)$:= the demand during $(0,t]$.

Note that $P\{\xi(t)\leq x\} = P\{\xi\leq x | T=t\}$. Analogously to (6.3.1) we have that

$$s(x,t) = \lambda E[\bar{T}(x,t)] + P\{\xi(t)>x\}, \quad x > 0. \tag{6.3.2}$$

Therefore we need an expression for $E[\bar{T}(x,t)]$.

We observe that the inventory levels immediately after an arrival correspond to a renewal process induced by F, the demand distribution. Then it follows that

$$E[T(x)] = \frac{M(x)}{\lambda}, \quad x > 0,$$

with

$$M(x) := \sum_{n=0}^{\infty} F^{n*}(x), \quad x \geq 0,$$

the renewal function associated with F (cf. also the analysis in section 1.3). On the other hand we have

$$E[T(x)] = \int_0^x E[T(x)|\xi(t)=y]dF_{\xi(t)}(y) + \int_x^{\infty} E[T(x)|\xi(t)=y]dF_{\xi(t)}(y)$$

with $F_{\xi(t)}(x) = P\{\xi(t)\leq x\}$. Analogously we obtain

$$E[\bar{T}(x,t)] = \int_x^{\infty} E[\bar{T}(x,t)|\xi(t)=y]dF_{\xi(t)}(y).$$

Here we used that $E[\bar{T}(x,t)|\xi(t)=y]=0$ for $0\leq y\leq x$. It follows from the definitions of $T(x)$ and $\bar{T}(x,t)$ that

$$E[T(x)|\xi(t)=y] + E[\bar{T}(x,t)|\xi(t)=y] = t, \quad y > x.$$

From the lack of memory of the Poisson arrival process we find

$$E[T(x)|\xi(t)=y] = t + E[T(x-y)], \quad 0 \leq y \leq x.$$

Combining the above relations we obtain

$$E[\bar{T}(x,t)] = t - \frac{M(x)}{\lambda} + \int_0^x \frac{M(x-y)}{\lambda} dF_{\xi(t)}(y), \quad x > 0.$$

Thus it follows from (6.3.2) that

$$s(x,t) = \lambda t - M(x) + \int_0^x M(x-y)dF_{\xi(t)}(y) + 1-F_{\xi(t)}(x), \quad x > 0. \tag{6.3.3}$$

Next we derive an expression for $c(x,t)$. Define the random variable $C(x,t)$ by

$$C(x,t) := \text{the cumulative backlog at epoch } t, \text{ given } X(0)=x.$$

By definition we have

$$C(x,t) = \int_0^t - X(s)1_{\{X(s)<0\}}ds.$$

This can be rewritten as

$$C(x,t) = \int_0^t (X(0)-X(s))ds - xt + \int_0^t X(s)1_{\{X(s)>0\}}ds.$$

Using $c(x,t) = E[C(x,t)]$ we obtain

$$c(x,t) = E[\int_0^t (X(0)-X(s))ds]-xt+E[\int_0^t X(s)1_{\{X(s)>0\}}ds]. \tag{6.3.4}$$

Let us first consider the first term on the right-hand side of (6.3.4). It is easy to see that this term is only a function of t. Therefore we define

$$k(t) := E[\int_0^t (X(0)-X(s))ds], \quad t \geq 0.$$

We can interprete $k(t)$ as follows. Assume that at epoch s a cost is incurred at rate $x-z$ if $X(s)=z$. Then $k(t)$ equals the expected cost incurred in $(0,t]$. Conditioning on the first interarrival time and noting that the cost incurred up to that time is zero, we obtain

$$k(t) = \int_0^t \{E[D](t-s)+k(t-s)\}\lambda e^{-\lambda s}ds, \quad t \geq 0,$$

which can be rewritten as

$$(6.3.5) \qquad k(t) = E[D][t - \frac{1}{\lambda}(1-e^{-\lambda t})] + \int_0^t k(s)\lambda e^{-\lambda(t-s)}ds.$$

Differentiation of equation (6.3.5) leads to

$$(6.3.6) \qquad k'(t) = E[D](1-e^{-\lambda t})+\lambda k(t) - \lambda \int_0^t k(s)\lambda e^{-\lambda(t-s)}ds.$$

Substitution of (6.3.5) into (6.3.6) yields

$$k'(t) = \lambda E[D]t, \quad t \geq 0.$$

Using $k(0)=0$ we thus find

$$(6.3.7) \qquad k(t) = \frac{\lambda E[D]t^2}{2}.$$

Next we want to find an expression for the last term in (6.3.4). This term depends on both x and t. Therefore we define for all x and all $t \geq 0$

$$K(x,t) := \int_0^t X(s)1_{\{X(s)>0\}}ds.$$

It is obvious that

$$(6.3.8) \qquad K(x,t) = 0, \quad x \leq 0, \ t \geq 0.$$

So let $x>0$. Assume that at epoch s a cost at rate z is incurred if $X(s)=z$ with $z>0$. Otherwise no cost is incurred at epoch s. Then $K(x,t)$ can be interpreted as the cost incurred in $(0,t]$. Analogously to the derivation of an expression for $s(x,t)$ we condition on the demand in the setup time t. This yields

$$E[K(x,t)] = \int_0^x E[K(x,t)|\xi(t)=y]dF_{\xi(t)}(y) +$$

$$+ \int_x^\infty E[K(x,t)|\xi(t)=y]dF_{\xi(t)}(y).$$

By the same conditioning arguments we have,

$$E[K(x,T(x))] = \int_0^x E[K(x,T(x))|\xi(t)=y]dF_{\xi(t)}(y)$$

$$+ \int_x^\infty E[K(x,T(x))|\xi(t)=y]dF_{\xi(t)}(y).$$

An expression for $E[K(x,T(x))]$ is given by equation (5.2.6) with h=1. However, for our present purpose it is more convenient to use the following expression for $E[K(x,T(x))]$,

$$E[K(x,T(x))] = \int_0^x \frac{(x-y)}{\lambda} dM(y), \quad x \geq 0.$$

This expression is derived as follows. Conditioning on the first arrival after epoch 0, we obtain

$$E[K(x,T(x))] = \frac{x}{\lambda} + \int_0^x E[K(x-y,T(x-y))]dF(y), \quad x \geq 0.$$

Clearly, the above expression for $E[K(x,T(x))]$ is the unique solution to this renewal equation.

Using the lack of memory of the Poisson arrival process, we obtain

$$E[K(x,T(x))|\xi(t)=y] = E[K(x,t)|\xi(t)=y]+E[K(x-y,T(x-y))],$$

$$0 \leq y \leq x.$$

Combining the above relations with $E[K(x,T(x))|\xi(t)=y]=E[K(x,t)|\xi(t)=y]$ for all $y>x$, we find

$$E[K(x,t)] = \int_0^x \frac{(x-y)}{\lambda} dM(y) - \int_0^x \int_0^{x-y} \frac{(x-y-z)}{\lambda} dM(y)dF_{\xi(t)}(y). \tag{6.3.9}$$

This expression for $E[K(x,t)]$ is also of interest with respect to the evaluation of the holding costs when these costs are linear in the stock on hand. Equations (6.3.4) and (6.3.7)-(6.3.9) together yield

$$c(x,t) = \begin{cases} \frac{\lambda E[D]t^2}{2} - xt + \int_0^x \frac{(x-y)}{\lambda} dM(y) - \\ \int_0^x \int_0^{x-y} \frac{(x-y-z)}{\lambda} dM(z)dF_{\xi(t)}(y), & x > 0. \\ \frac{\lambda E[D]t^2}{2} - xt, & x \leq 0 \end{cases} \tag{6.3.10}$$

Now we are in a position to give exact expressions for $s(T)$ and $c(T)$. Using the definitions of $s(x,t)$ and $c(x,t)$ and the memorylessness of the exponential interarrival times we obtain

$$s(T) = \int_0^\infty \int_0^\infty s(m-u,t)\,d_u(1-p(M-m,u))\,dF(t)+p(M-m,m),$$

$$c(T) = \int_0^\infty \int_0^\infty c(m-u,t)\,d_u(1-p(M-m,u))\,dF(t),$$

with $F(t) = P\{T\leq t\}$. Substituting (6.3.1), (6.3.3) and (6.3.10) into the equations above and using the definitions of $F_{\xi(t)}$, F_ξ and F_η it follows after some algebra that

$$(6.3.11)\quad s(T) = \lambda E[T] - \int_0^m M(m-u)\,d_u(1-p(M-m,u)) + \int_0^m \int_0^{m-y} M(m-y-u)\,d_u(1-p(M-m,u))\,dF_\xi(y)+1-F_\eta(m).$$

$$(6.3.12)\quad c(T) = \frac{\lambda E[D]}{2} E[T^2] - (m-E[U(M-m)])E[T] + \int_0^m \int_0^{m-u} \frac{M(m-u-z)}{\lambda}\,dz\,d_u(1-p(M-m,u)) - \int_0^m \int_0^{m-y} \int_0^{m-y-u} \frac{M(m-y-u-z)}{\lambda}\,dz\,d_u(1-p(M-m,u))\,dF_\xi(y).$$

From equation (1.3.34) we know that for the case of M=m

$$p(M-m,u) = 1-F(u), \qquad u \geq 0.$$

From the definition of M(x) we have that

$$\int_0^x M(x-y)\,dF(y) = M(x)-1, \qquad x \geq 0.$$

$$\int_0^x \int_0^{x-y} M(x-y-z)\,dz\,dF(y) = \int_0^x M(y)\,dy-x, \qquad x \geq 0.$$

Substituting these results into (6.3.11) and (6.3.12), we obtain for the case of M=m

$$(6.3.13)\quad s(T) = \lambda E[T]-M(m)+\int_0^m M(m-y)\,dF_\xi(y)+1-F_\xi(m)+1-F_\eta(m).$$

$$(6.3.14)\qquad c(T) = \frac{\lambda E[D]}{2} E[T^2] - (m-E[D])E[T] + \int_0^m \frac{M(y)}{\lambda}\, dy - \frac{m}{\lambda} - \int_0^m \int_0^{m-y} \frac{M(z)}{\lambda}\, dz dF_\xi(y) + \int_0^m \frac{(m-y)}{\lambda}\, dF_\xi(y).$$

When M-m satisfies condition 1.3.1 we use approximation 1.3.2,

$$p(M-m,u) \cong \frac{1}{E[D]} \int_u^\infty (1-F(y))dy, \qquad u \geq 0.$$

It can easily be verified that

$$\int_0^x M(x-y)(1-F(y))dy = x, \qquad x \geq 0$$

$$\int_0^x \int_0^{x-y} M(x-y-z)dz(1-F(y))dy = \frac{x^2}{2}, \qquad x \geq 0.$$

Again we substitute these results into (6.3.11) and (6.3.12) to obtain for the case of M-m satisfying condition 1.3.1

$$(6.3.15)\qquad s(T) \cong \lambda E[T] - \frac{m}{E[D]} + \int_0^m \frac{(m-y)}{E[D]}\, dF_\xi(y) + 1-F_\eta(m)$$

$$(6.3.16)\qquad c(T) \cong \frac{\lambda E[D]E[T^2]}{2} - (m - \frac{E[D^2]}{2E[D]})\, E[T] + \frac{m^2}{2\lambda E[D]} - \int_0^m \frac{(m-y)^2}{2\lambda E[D]}\, dF_\xi(y).$$

Hence for the case of M-m satisfying condition 1.3.1 equations (6.3.15) and (6.3.16) in combination with $\hat{F}_\xi(x)$ yield tractable expressions for $s(T)$ and $c(T)$. However, for the case of M=m we need to have an expression for M(x). In general no computationally tractable expression for M(x) is available. Therefore we resort to an approximation of a simple form. From renewal theory we know that

$$(6.3.17)\qquad \lim_{x\to\infty} [M(x) - (\frac{x}{E[D]} + \frac{E[D^2]}{2(E[D])^2})] = 0.$$

Assuming that F has a density we obtain

$$M'(0) = F'(0),$$

where we used the fact that F(0)=0. In view of these boundary conditions we suggest the following approximation $\hat{M}(x)$ for M(x).

$$\hat{M}(x) := 1 + \frac{x}{E[D]} + \frac{(c_D^2-1)}{2}\left(1-e^{-2\left(\frac{F'(0)-1/E[D]}{c_D^2-1}\right)x}\right), \quad x \geq 0$$

As usual c_D^2 denotes the squared coefficient of variation of D. If F is a K_2-distribution then $\hat{M}(x)=M(x)$. In order to have that (6.3.17) holds with M(x) replaced by $\hat{M}(x)$, a necessary condition is that

$$(6.3.18) \qquad c_D^2 \neq 1 \Rightarrow \frac{F'(0)-1/E[D]}{c_D^2-1} > 0.$$

A sufficient condition for (6.3.18) to hold is that F is NBUE (NWUE) and $F'(0)\neq 1/E[D]$. This condition is satisfied for gamma distribution functions and for mixtures of Erlang-k and Erlang-(k-1) distributions with the same scale parameter, provided $c_D^2 \leqq 1$. For a more detailed discussion of the ideas that lead to the approximation for M(x) we refer to section (7.4).

Finally, a combination of (6.3.13)-(6.3.16) with the definitions of $\hat{F}_\xi(x)$, $\hat{F}_\eta(x)$ and $\hat{M}(x)$ yields the following approximations.

Approximation 6.3.1.

$$s(T) \cong \begin{cases} \lambda E[T]-\hat{M}(m)+\int_0^m \hat{M}(m-y)d\hat{F}_\xi(y)+1-\hat{F}_\xi(m)+1-\hat{F}_\eta(m), & M = m \\ \lambda E[T]-\frac{m}{E[D]}+\int_0^m \frac{(m-y)}{E[D]}\,d\hat{F}_\xi(y)+1-\hat{F}_\eta(m) & M-m \geq \Delta_0. \end{cases}$$

Approximation 6.3.2.

$$c(T) \cong \begin{cases} \frac{\lambda E[D]}{2}E[T^2]-(m-E[D])E[T]+\int_0^m \frac{\hat{M}(y)}{\lambda}\,dy-\frac{m}{\lambda} \\ \qquad -\int_0^m\int_0^{m-y}\frac{\hat{M}(z)}{\lambda}\,dz\,d\hat{F}_\xi(y)+\int_0^m \frac{(m-y)}{\lambda}\,d\hat{F}_\xi(y), & M = m \\ \frac{\lambda E[D]}{2}E[T^2]-(m-\frac{E[D^2]}{2E[D]})E[T]+\frac{m^2}{2\lambda E[D]}-\int_0^m \frac{(m-y)^2}{2\lambda E[D]}\,d\hat{F}_\xi(y), \\ & M-m \geq \Delta_0 \end{cases}$$

Here the number Δ_0 is given by

$$\Delta_0 = \begin{cases} E[D], & c_D^2 \leq 1 \\ 1\tfrac{1}{2}\, c_D^2 \cdot E[D], & c_D^2 > 1 \end{cases} \tag{6.3.19}$$

These approximations together with those already obtained in chapter 1 and section 6.2 enable us to compute approximations for $E[T_B]$, $E[B_B]$, $E[Q_B]$, $E[S_B]$ and $E[C_B]$ from equations (6.1.7)-(6.1.11). In this way we obtain approximations for the service measures under a given (m,M)-rule. In section 6.5 we discuss the accuracy of these approximations with some sensitivity analysis.

6.4. Average holding and setup costs.

In the previous sections we have provided the tools to obtain approximate expressions for the various service measures. We assumed that both m and M were given. It is an important problem to determine an (m,M)-rule that minimizes long-run average costs subject to some given service level constraint. We will approximately solve this problem for the following cost structure. Holding costs are incurred at a rate h.x if the on-hand inventory equals x>0, otherwise no holding costs are incurred. A fixed setup cost K≧0 is incurred each time the production facility is reactivated.

We define the random variable C_h by

C_h := the holding cost incurred during a regeneration cycle.

Then it follows from the theory of regenerative processes that

$$\frac{E[C_h]}{E[T]} = \text{the long-run average holding costs per unit time} \tag{6.4.1}$$

$$\frac{K}{E[T]} = \text{the long-run average setup costs per unit time.} \tag{6.4.2}$$

It remains to determine an expression for $E[C_h]$. Again it follows from the lack of memory of the Poisson arrival process and $\pi_2 > \lambda E[D]$ that $E[C_h]$ is the same for both the lost-sales and backlog model. So let us consider the backlog model.

We recall the following definitions given in chapter 5.

$k_1(x,m)$:= the expected holding cost incurred until the inventory level drops below m, given that at epoch 0 the inventory level equals x+m, $x \geq 0$, and production is off.

$k_2(x)$:= the expected holding cost incurred until the inventory reaches the value M, given that at epoch 0 the inventory level equals $x \leq M$ and production is on.

We also define for a given (m,M)-rule

$k(T)$:= the expected holding cost incurred during the setup time.

Then it is immediately clear that

$$E[C_h] = k_1(M-m,m) + k(T) + \int_0^\infty k_2(m-y)dF_\eta(y). \tag{6.4.3}$$

It follows from equations (1.3.35), (5.1.3), (5.1.5), (5.2.6), (5.5.1) and approximations 1.3.3 and 5.1.1 that

$$k_1(M-m,m) \cong \frac{h}{\lambda E[D]}\left[\frac{(M-m)^2}{2} - \frac{E[U^2]}{2} + \left(\frac{E[D^2]}{2E[D]} + m\right)(M-m+E[U])\right], \tag{6.4.4}$$

with $E[U]$ and $E[U^2]$ being given by

$$E[U] = \begin{cases} E[D] & \text{when } M = m \\ \dfrac{E[D^2]}{2E[D]} & \text{when } M-m \geq \Delta_0, \end{cases} \tag{6.4.5}$$

$$E[U^2] = \begin{cases} E[D^2] & \text{when } M = m \\ \dfrac{E[D^3]}{3E[D]} & \text{when } M-m \geq \Delta_0, \end{cases} \tag{6.4.6}$$

where Δ_0 is given by (6.3.19).
Equation (5.4.7) gives an exact expression for $k_2(x)$,

$$k_2(x) = h\left\{\frac{M^2-x^2}{2(\pi_2-\lambda E[D])} - \frac{\lambda E[D^2]}{2(\pi_2-\lambda E[D])^2}(M-x) + c_B(x)\right\}, \quad x \leq M. \tag{6.4.7}$$

An expression for $k(T)$ has in fact been obtained already in the derivation of an expression for $c(T)$. By the definition of $K(x,t)$ given in section 6.3, we have

$$(6.4.8) \qquad k(T) = h \int_0^\infty \int_0^\infty E[K(m-u,t)]\, d_u(1-p(M,m,u))dF(t).$$

Proceeding as in the derivation of the approximations 6.3.1 and 6.3.2 for $s(T)$ and $c(T)$, respectively, we obtain from (6.4.8) and the approximation for $p(M-m,u)$ for the case of $M-m \geq \Delta_0$

$$(6.4.9) \qquad k(T) \cong \begin{cases} h[\int_0^m \frac{M(y)}{\lambda}\, dy - \frac{m}{\lambda} - \int_0^m \int_0^{m-y} \frac{M(z)}{\lambda}\, dz dF_\xi(y) \\ \qquad + \int_0^m \frac{(m-y)}{\lambda}\, dF_\xi(y)], \quad M = m \\ h[\frac{m^2}{2\lambda E[D]} - \int_0^m \frac{(m-y)^2}{2\lambda E[D]}\, dF_\xi(y)], \quad M-m \geq \Delta_0 \end{cases}$$

The approximation for $k(T)$ given by (6.4.9) is exact for the case of $M=m$.

From approximation 1.4.3 and equation (1.4.36) we can find an approximation for $c_B(x)$. Let $\tilde{c}_B(x)$ denote this approximation for all $x \leq M$. Then we define

$$E_{app}[C] := \int_0^\infty \tilde{c}_B(m-y)d\hat{F}_\eta(y).$$

Some reflections reveal that $E_{app}[C]$ approximates the difference between the cumulative backlog at epoch T and the cumulative backlog at the earlier epoch at which the production facility starts producing. To avoid lengthy expressions we do not elaborate the expression for $E_{app}[C]$. It can be easily derived from equation (1.4.36), approximation 1.4.3 and the approximation for $\hat{F}_\eta(x)$. Combining (6.4.3), (6.4.4), (6.4.7) and (6.4.9) with the approximations given in section 6.3 for $F_\eta(x)$, $F_\xi(x)$ and $M(x)$ we obtain the following approximation for $E[C_h]$.

Approximation 6.4.1.

$$E[C_h] \cong \begin{cases} h\{\int_0^m \frac{\hat{M}(y)}{\lambda}\,dy - \int_0^m \int_0^{m-y} \frac{\hat{M}(z)}{\lambda}\,dz d\hat{F}_\xi(y) + \int_0^m \frac{(m-y)}{\lambda}\,d\hat{F}_\xi(y) \\ + (m - \frac{\lambda E[D^2]}{2(\pi_2-\lambda E[D])})\frac{E[\eta]}{(\pi_2-\lambda E[D])} - \frac{E[\eta^2]}{2(\pi_2-\lambda E[D])} + E_{app}[C]\} \\ \qquad\qquad \text{when } M = m \\ h\{\frac{\pi_2}{\lambda E[D](\pi_2-\lambda E[D])}\ [\frac{(M-m)^2}{2} - \frac{E[\eta^2]}{2} + m(M-m+E[\eta])] \\ + E_{app}[C] + \int_m^\infty \frac{(m-y)^2}{2\lambda E[D]}\,d\hat{F}_\xi(y) \\ + \frac{\lambda E[D^2]}{2}\,(M-m+E[\eta])[\frac{1}{(\lambda E[D])^2} - \frac{1}{(\pi_2-\lambda E[D])^2}]\} \\ \qquad\qquad \text{when } M-m \geq \Delta_0. \end{cases}$$

Next we define

$g(\Delta,m)$:= the long-run average costs per unit time if the inventory is controlled by an $(m,m+\Delta)$-rule.

Then it is obvious that

$$g(\Delta,m) = \frac{E[C_h]+K}{E[T]}\ .$$

Approximation 6.4.1 provides a tractable expression for $E[C_h]$, while an expression for $E[T]$ follows from equations (1.4.1), (6.1.7), (6.1.17) and approximation 1.4.2. Note that

$$E[T_B] = \frac{\pi_2(\Delta+E[\eta])}{\lambda E[D](\pi_2-\lambda E[D])}\ . \tag{6.4.10}$$

Let

$r(\Delta,m)$:= the value of any service measure if the inventory is controlled by an $(m,m+\Delta)$-rule,

where we restrict to the α-, β-, γ- and δ-service measures. Then we can approximately solve the following problem.

Problem Pb . Find (Δ^*,m^*) such that

$$g(\Delta^*,m^*) = \min\{g(\Delta,m) \mid \Delta\geq 0,\ m\geq 0,\ r(\Delta,m)=\alpha\},$$

where α is the prespecified level of the service measure under consideration.

We approximately solve problem Pb by the method described below Pb 1 in section 5.5.

In section 5.5 we argued that for the production-inventory model with zero setup time the optimal difference Δ^* is insensitive to the service level constraint provided that the required service level is sufficiently high. It is important to note that in the present model with a positive setup time this phenomenon occurs as well. However, it turns out that the longer the setup time, the higher the required service level should be before the optimal difference Δ^* becomes constant. On the other hand we find that the longer the setup time, the smaller the optimal difference Δ^* becomes. To gain insight in these empirical findings we take a closer look at approximation 6.4.1 for $\Delta\geq\Delta_0$.

Comparing approximation 6.4.1 with approximation 5.5.1 we observe that the former approximation can be obtained from the latter by substituting η for U and adding the integral with respect to $F_\xi(x)$. Then we would like to have a similar result as approximation 5.5.4, which holds for high service requirements. Therefore we proceed along the same lines as in section 5.5.

We recall that to apply the approximations we must have that for all $x\geq 0$

$$1-F(x) \leq Ce^{-\kappa x}. \tag{6.4.11}$$

Without loss of generality we assume that $\kappa>\delta$, where δ is defined by (1.4.7). It follows from (6.4.11) that for all positive ε,

$$\tilde{F}(-\kappa+\varepsilon) < \infty,$$

where $\tilde{F}(s)$ is the Laplace-Stieltjes transform of F(x). Then it is easily established that for all $x\geq 0$ and $n\geq 1$

$$(6.4.12) \qquad 1-F^{n*}(x) \leq C[1+\tilde{F}(-\kappa+\varepsilon)]^{n-1}e^{-(\kappa-\varepsilon)x}, \quad \varepsilon > 0.$$

Let $\tilde{p}(M-m,s)$, $\tilde{F}(s)$, $\tilde{F}_{\xi}(s)$, $\tilde{F}_{\eta}(s)$ denote the Laplace-Stieltjes transforms of $p(M-m,u)$, $F(t)$, $F_{\xi}(x)$ and $F_{\eta}(x)$, respectively. It follows from the definition of $F(t)$ and the fact that the demand process is a compound Poisson process that

$$1-F_{\xi}(x) = \int_0^{\infty} \{ \sum_{n=1}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} (1-F^{n*}(x))\} dF(t)$$

and thus (6.4.12) yields for all $x \geq 0$

$$(6.4.13) \qquad 1-F_{\xi}(x) \leq \frac{C}{1+\tilde{F}(-\kappa+\varepsilon)} [\tilde{F}(-\lambda\tilde{F}(-\kappa+\varepsilon))-\tilde{F}(\lambda)]e^{-(\kappa-\varepsilon)x}, \quad \varepsilon > 0.$$

Now it is crucial whether or not $\tilde{F}(-\lambda\tilde{F}(-\kappa+\varepsilon))$ is finite for some positive ε with $\delta<\kappa-\varepsilon$. If not then (6.4.13) has no meaning and it will follow from the analysis below that we may not expect that Δ^* converges to some constant as the service requirement increases. If $\tilde{F}(-\lambda\tilde{F}(-\kappa+\varepsilon))$ is finite for some positive ε with $\delta<\kappa-\varepsilon$ then $F_{\xi}(x)$ has an exponential tail and, moreover, $\lim_{x\to\infty} e^{\delta x}(1-F_{\xi}(x))=0$. We note that $\tilde{F}(-\lambda\tilde{F}(-\kappa+\varepsilon))$ is finite when $F(t)$ has a finite support.

From now on we assume that there exists a positive ε_0 such that

$$(6.4.14) \qquad \tilde{F}(-\lambda\tilde{F}(-\kappa+\varepsilon_0)) < \infty \text{ and } \delta < \kappa-\varepsilon_0.$$

It should be noted that this assumption is rather restrictive. For example, if T is exponential then (6.4.14) will be violated as $E[T]$ increases beyond $1/(\lambda\tilde{F}(-\delta))$. Equation (5.5.9) states that

$$(6.4.15) \qquad p(\Delta,u) \leq C'e^{-\kappa u}, \quad u \geq 0.$$

Since $\eta=\xi+U(\Delta)$ it follows from (6.4.13)-(6.4.15) that

$$(6.4.16) \qquad 1-F_{\eta}(x) \leq \{C' + \frac{C\tilde{p}(\Delta,-\kappa+\varepsilon_0)}{1-\tilde{F}(-\kappa+\varepsilon_0)} [\tilde{F}(-\lambda F(-\kappa+\varepsilon_0))-\tilde{F}(\lambda)]\}^{-(\kappa-\varepsilon_0)x}.$$

Proceeding along the same lines as in the derivation of equation (5.5.13), we obtain from (6.4.13) and (6.4.16),

$$(6.4.17)\qquad \lim_{m\to\infty} e^{\delta m} \int_0^\infty c_B(m-y)\,dF_\eta(y) = \frac{1}{\pi_2\delta^3\nu}\,[\tilde{F}_\eta(-\delta)-e^{-\delta\Delta}]$$

$$(6.4.18)\qquad \lim_{m\to\infty} e^{\delta m} \int_m^\infty (m-y)^2\,dF_\xi(y) = 0$$

$$(6.4.19)\qquad \lim_{m\to\infty} e^{\delta m} E[B_L] = \frac{\pi_2-\lambda E[D]}{\pi_2\delta^2\nu}\,[\tilde{F}_\eta(-\delta)-e^{-\delta\Delta}]$$

$$(6.4.20)\qquad \lim_{m\to\infty} e^{\delta m}(1-r(\Delta,m))E[T] = c_r[\tilde{F}_\eta(-\delta)-e^{-\delta\Delta}],$$

where $r(\Delta,m)$ is the level of the service measure considered and c_r is some positive constant also depending on this service measure.

It follows from the expressions for $p(\Delta,u)$ and $\eta=\xi+U(\Delta)$ that

$$\tilde{F}_\eta(s) \cong \begin{cases} \tilde{F}(s)\tilde{F}_\xi(s) & \text{when } \Delta = 0 \\ \frac{1-\tilde{F}(s)}{sE[D]}\cdot\tilde{F}_\xi(s) & \text{when } \Delta \geq \Delta_0 \end{cases}$$

We already derived an expression for $1-F_\xi(x)$. Using this expression we find an expression for $\tilde{F}_\xi(s)$. It is readily seen that

$$\tilde{F}_\eta(s) \cong \begin{cases} \tilde{F}(s)\tilde{F}(\lambda(1-\tilde{F}(s))) & \text{when } \Delta = 0 \\ \frac{1-\tilde{F}(s)}{sE[D]}\,\tilde{F}(\lambda(1-\tilde{F}(s))) & \text{when } \Delta \geq \Delta_0 \end{cases}$$

Using the definition of δ given by equation (1.4.7), we obtain

$$(6.4.21)\qquad \tilde{F}_\eta(-\delta) \cong \begin{cases} (1+\frac{\pi_2\delta}{\lambda})\tilde{F}(-\pi_2\delta) & \text{when } \Delta = 0 \\ \frac{\pi_2}{\lambda E[D]}\,\tilde{F}(-\pi_2\delta) & \text{when } \Delta \geq \Delta_0 \end{cases}$$

Note that $\tilde{F}(-\pi_2\delta)$ is finite because of (6.4.14). Combining approximation 6.4.1 and equations (6.4.17)-(6.4.19) and (6.4.21), we find

Approximation 6.4.2: For m sufficiently large and $\Delta\geq\Delta_0$, we have

$$g(\Delta,m) \cong \tilde{g}(\Delta,m)$$

with

$$\tilde{g}(\Delta,m) := [K+h\{\frac{\pi_2}{\lambda E[D](\pi_2-\lambda E[D])}(\frac{\Delta^2}{2} - \frac{E[\eta^2]}{2} + m(\Delta+E[\eta]))$$

$$\frac{e^{-\delta m}}{\pi_2\delta^3\nu}(c_\delta(T)-e^{-\delta\Delta}) +$$

$$+ \frac{\lambda E[D^2]}{2}(\Delta+E[\eta])(\frac{1}{(\lambda E[D])^2} - \frac{1}{(\pi_2-\lambda E[D])^2})\}]$$

$$\times[\frac{\pi_2(\Delta+E[\eta])}{\lambda E[D](\pi_2-\lambda E[D])} - c_\ell\frac{(c_\delta(T)-e^{-\delta\Delta})}{(\pi_2-\lambda E[D])}e^{-\delta m}]-1$$

and where

$$c_\delta(T) = \frac{\pi_2}{\lambda E[D]}\tilde{F}(-\pi_2\delta),$$

$$c_\ell = \begin{cases} 0 & \text{for the backlog model} \\ \dfrac{\pi_2-\lambda E[D]}{\pi_2\delta^2\nu} & \text{for the lost-sales model} \end{cases}$$

We emphasize the fact that $c_\delta(T)$ depends on the distribution of T.

Let $r(\Delta,m)$ be the service level of one of the service measures in this chapter. Define

$$\tilde{r}(\Delta,m) := 1-c_r(c_\delta(T)-e^{-\delta\Delta})e^{-\delta m}[\frac{\pi_2(\Delta+E[\eta])}{\lambda E[D](\pi_2-\lambda E[D])} - c_\ell\frac{(c_\delta(T)-e^{-\delta\Delta})}{(\pi_2-\lambda E[D])}e^{-\delta m}]-1$$

where c_r is given through (6.4.20). We define the following approximate version of problem Pb.

Problem $\tilde{Pb}$. Find $(\tilde{\Delta}^*,\tilde{m}^*)$ such that

$$g(\tilde{\Delta}^*,\tilde{m}^*) = \min\{\tilde{g}(\Delta,m) \mid \Delta\geq 0,\ m\geq 0,\ \tilde{r}(\Delta,m)=\alpha\}$$

where α is the prespecified level of the service measure considered.

As in section 5.5 we find that $\tilde{\Delta}^*$ is a positive root of

$$z(\Delta) = 0,$$

where $z(\Delta)$ is defined by

$$z(\Delta) := h\Delta(\Delta+E[\eta]) + \frac{h(\Delta+E[\eta])^2}{c_\delta(T)-e^{-\delta\Delta}} e^{-\delta\Delta} - \frac{h(\Delta^2-E[\eta^2])}{2}$$

$$- \frac{h(\Delta+E[\eta])}{\delta} - \frac{K\lambda E[D](\pi_2-\lambda E[D])}{\pi_2} .$$

As in section 5.5 the function $z(\Delta)$ does not depend on the service measure and the specific model. To obtain the value of $\tilde{\Delta}^*$ for a given set of model parameters one can use the algorithm described below (5.5.17). We emphasize that the function $z(\Delta)$ is meaningful only when $\Delta \geq \Delta_0$, since only for these values of Δ our approximations are valid.

In the next section we present numerical results. Anticipating on these results we make some qualitative statements about the relation between the switchtime T and the convergence of Δ^* to $\tilde{\Delta}^*$ as the required service level tends to 1. It follows from (6.4.13) and (6.4.18) that $1-F_\xi(x) \leq C_\xi e^{-(\kappa-\varepsilon_0)x}$, where

$$C_\xi := \frac{C}{1+\tilde{F}(-\kappa+\varepsilon_0)} [\tilde{F}(-\lambda\tilde{F}(-\kappa+\varepsilon_0))-\tilde{F}(\lambda)].$$

However, the values of x for which this exponential behaviour manifests will depend on the magnitude of C_ξ. One typically sees that the larger C_ξ, the larger the values of x for which $1-F_\xi(x)$ decreases exponentially.

Let us assume that $T_1 \overset{st}{\leq} T_2$ where "$\overset{st}{\leq}$" denotes "stochastically smaller", i.e.

$$X_1 \overset{st}{\leq} X_2 \Leftrightarrow P\{X_1>x\} \leq P\{X_2>x\} \quad \forall\, x \geq 0 .$$

Also, let $\tilde{F}_1$, $\tilde{F}_2$ denote the Laplace-Stieltjes transform of T_1, T_2. Then we have the following result,

$$T_1 \overset{st}{\leq} T_2 \Rightarrow \tilde{F}_1(s) \leq \tilde{F}_2(s) \text{ for all } s < 0 \text{ with } \tilde{F}_2(s) < \infty.$$

Then it is immediately clear that C_ξ increases as T stochastically increases. This indicates that as T gets stochastically larger, it takes "more time" before the tail of $F_\xi(x)$ becomes exponential. Equation (6.4.16) implies that the same holds for the tail of $F_\eta(x)$. Thus as T gets larger

the convergence in equations (6.4.17)-(6.4.20) gets slower. Hence Δ^* converges slower to Δ^* as T increases stochastically. We empirically verified this claim for the important case of a deterministic setup time T_D. We used the method described below problem Pb 1 in section 5.5 to obtain the value of Δ^* and solved the equation $z(\Delta)=0$ to obtain Δ^*. Another noteworthy finding from this numerical experience is that as the value of T_D increases, the value of Δ^* decreases, where we keep the required service level fixed.

For some value of T_D, say T_0 we find $\tilde{\Delta}^*=\Delta_0$. If T_D increases beyond T_0, then $\tilde{\Delta}^*<\Delta_0$. However, when $\tilde{\Delta}^*<\Delta_0$ then our approximations are no longer accurate. Hence we may not expect that the true Δ^* converges to $\tilde{\Delta}^*$. In this case we suggest to set $\tilde{\Delta}^*$ equal to zero or to Δ_0, depending on what gives lowest costs. This suggestion is based on the empirically verified result that the cost function is usually very flat around its minimum (cf. Peterson and Silver [1979]).

In the next section we provide additional support to the above statements by presenting numerical results. We draw conclusions with respect to the accuracy of the approximations. Also, we make further comments on the sensitivity of the (m,M)-rules to the underlying demand distribution and to the value of T_D.

6.5. Numerical results and conclusions.

In this section we present numerical results to get some idea of the performance of the approximations for the case of a positive setup time. In this case we need also two-moment approximations for the distribution functions $F_\xi(x)$ and $F_\eta(x)$ and the renewal function $M(x)$, and thus we may not expect that the resulting approximations show the same good performance as in the case of no setup time. Nevertheless, it turns out that the quality of the approximations is good enough for practical purposes.

Table 6.5.1 deals with the approximate (m,M)-rules for the γ-service level in the backlog model. We have chosen to consider the γ-service level since the computation of the approximate γ-service level involves most of the approximations. In all examples $\lambda=1$, $E[D]=1$ and we assume a deterministic set-up time T equal to 2. The value of M-m is predetermined by formula (1.5.1) with h=1 and where K has the two values 0 and 25. For the case of K=0 we find an (m,M)-rule with m=M. The production rate π_2 has the three

values 1.25, 2 and 5. The required service level γ is varied as 0.95 and 0.99. The coefficient of variation of the demand size D, c_D^2, runs through the values 0, 1/3, 2/3 and 2. We consider the following demand distributions

(i) deterministic demand ($c_D^2=0$).
(ii) Erlang-3 demand ($c_D^2=1/3$).
(iii) mixture of Erlang-1 and Erlang-2 demand with the same scale parameters ($c_D^2=2/3$).
(iv) hyperexponential demand of order two with balanced means ($c_D^2=2$).

For deterministic demand we used the exact undershoot distribution The resulting approximation to $s(T)$ for the case of $M \geq m$ is similar to the one given by approximation 6.3.1 for the case of M=m. For the other demand distributions the used approximations are as in the sections 6.2 and 6.3. The actual γ-service levels of the approximate (m,M)-rules have been determined by computer simulation. In each example we have simulated 250,000 individual demands. As before the notation 0.954(6) is used to denote that the simulated value is 0.954 with [0.948,0.960] as 95% confidence interval.

From the results in table 6.5.1 we draw the following conclusions. For the case of deterministic demand the performance of the approximations deteriorates as π_2 increases. This finding applies also to the other demand distributions when M=m. If $c_D^2>0$ and M>m, the numerical results show that the approximations perform very well for all values of π_2. For the other service measures we found similar results.

The sensitivity of the level m to the underlying demand distribution is considered in table 6.5.2. The approach to this problem is different from the one used in section 1.5 leading to the results given in table 1.5.3. There we computed the approximate (m,M)-rules for several demand distributions. Here we compute the approximate (m,M)-rule for a mixture of Erlang-k and Erlang-(k-1) demand distributions with the same scale parameters and for the resulting (m,M)-rule we determine the actual service levels for several other demand distributions. Thus the testing of the quality of a two-moment approximation is combined with sensitivity analysis.

Table 6.5.1. The approximate (m,M)-rules and their actual γ-service levels.

		$c_D^2=0$			$c_D^2=1/3$		
π_2		m	M	γ_{act}	m	M	γ_{act}
1.25	0.95	9.55	9.55	0.954(6)	11.91	11.91	0.951(6)
2	0.95	5.96	5.96	0.957(3)	6.81	6.81	0.953(3)
5	0.95	5.38	5.38	0.967(2)	5.86	5.86	0.956(2)
1.25	0.99	13.31	13.31	0.991(3)	17.09	17.09	0.992(3)
2	0.99	7.96	7.96	0.994(1)	9.22	9.22	0.992(2)
5	0.99	7.30	7.30	0.996(1)	8.04	8.04	0.993(1)
1.25	0.95	8.17	11.33	0.953(6)	10.25	13.41	0.952(6)
2	0.95	4.66	9.66	0.949(3)	4.97	9.97	0.950(3)
5	0.95	3.65	9.97	0.937(2)	3.87	10.20	0.950(2)
1.25	0.99	11.94	15.10	0.992(3)	15.42	18.59	0.990(4)
2	0.99	6.76	11.76	0.993(1)	7.53	12.53	0.992(2)
5	0.99	5.68	12.00	0.996(1)	6.23	12.55	0.991(1)
		$c_D^2=2/3$			$c_D^2=2$		
π_2		m	M	γ_{act}	m	M	γ_{act}
1.25	0.95	14.31	14.31	0.951(6)	24.49	24.49	0.955(8)
2	0.95	7.72	7.72	0.955(3)	11.25	11.25	0.946(5)
5	0.95	6.43	6.43	0.954(2)	7.87	7.87	0.941(2)
1.25	0.99	20.89	20.89	0.990(4)	37.47	37.47	0.989(5)
2	0.99	11.70	11.70	0.991(2)	17.25	17.25	0.988(2)
5	0.99	9.04	9.04	0.993(2)	12.45	12.45	0.986(1)
1.25	0.95	12.58	15.74	0.954(9)	22.79	25.95	0.950(10)
2	0.95	5.73	10.73	0.950(3)	9.24	14.24	0.951(4)
5	0.95	4.28	10.60	0.950(2)	6.04	12.37	0.949(2)
1.25	0.99	19.16	22.33	0.992(3)	35.78	38.94	0.988(5)
2	0.99	8.86	13.86	0.991(1)	15.29	20.29	0.989(2)
5	0.99	7.08	13.40	0.992(1)	10.84	17.17	0.990(2)

We deal with the β-service level for both the backlog model and the lost-sales model. In each example we have chosen $\lambda=1$, $E[D]=1$, $T=2$ and predetermined M-m by using formula (1.5.1) with $h=1$ and $K=25$. The production rate π_2 has the three values 1.25, 2 and 5, β is varied as 0.95 and 0.99 and c_D^2 has the two values 1/3 and 4/5. We do not consider values of c_D^2 that are larger than 1, since we already concluded in section 1.5 for zero setup time that when $c_D^2>1$ the switching level m is sensitive to the underlying demand distribution. The following three demand distributions are considered.

(i) Weibull distribution.

(ii) uniform distribution.

(iii) shifted exponential distribution.

Each of these distributions is completely determined by specifying the first two moments, where the uniform distribution requires that $0<c_D^2\leq 1/3$ and the shifted exponential distribution requires $0<c_D^2\leq 1$. The shifted exponential distribution corresponds to the sum of an exponential random variable and a positive constant.

We computed the approximate (m,M)-rules using Erlang-3 demand ($c_D^2=1/3$) and a mixture of Erlang-1 and Erlang-2 demand distributions with the same scale parameters ($c_D^2=4/5$). The actual β-service levels β_{Weib}, β_{unif} and β_{shif} have been determined by computer simulation, where in each example 200,000 customers are simulated.

Table 6.5.2. The actual β-service levels for various demand distributions.

backorders			$c_D^2=1/3$			$c_D^2=4/5$		
π_2	β	m_{app}	β_{Weib}	β_{unif}	β_{shif}	m_{app}	β_{Weib}	β_{shif}
1.25	0.95	9.90	0.951(5)	0.954(6)	0.948(5)	13.41	0.950(6)	0.949(6)
2	0.95	4.62	0.952(3)	0.954(3)	0.948(3)	5.91	0.949(4)	0.947(4)
5	0.95	3.52	0.950(2)	0.951(2)	0.950(2)	4.31	0.950(2)	0.949(2)
1.25	0.99	15.08	0.990(4)	0.991(4)	0.989(5)	20.56	0.990(5)	0.989(6)
2	0.99	7.17	0.992(2)	0.993(1)	0.990(2)	9.27	0.990(2)	0.989(2)
5	0.99	5.84	0.992(1)	0.993(1)	0.990(1)	7.26	0.990(1)	0.990(1)
lost-sales			$c_D^2=1/3$			$c_D^2=4/5$		
π_2	β	m_{app}	β_{Weib}	β_{unif}	β_{shif}	m_{app}	β_{Weib}	β_{shif}
1.25	0.95	5.21	0.949(2)	0.950(3)	0.948(3)	6.99	0.949(3)	0.949(3)
2	0.95	3.49	0.950(2)	0.951(2)	0.948(2)	4.45	0.950(2)	0.949(2)
5	0.95	3.19	0.950(2)	0.951(1)	0.949(2)	3.90	0.950(2)	0.949(2)
1.25	0.99	10.03	0.990(1)	0.990(2)	0.989(2)	13.58	0.990(2)	0.989(2)
2	0.99	6.12	0.991(1)	0.992(1)	0.990(1)	7.87	0.990(1)	0.990(2)
5	0.99	5.53	0.991(1)	0.992(1)	0.990(1)	6.87	0.990(1)	0.990(1)

From the results given in table 6.5.2 we conclude that the approximate (m,M)-rules perform quite well. This is in accordance with the conclusion drawn above with respect to the accuracy of the approximation for the case of $c_D^2>0$ and M-m>0. Also note that, as opposed to the pure-inventory models, the production-inventory backlog model and lost-sales model yield quite different results (cf. Tijms and Groenevelt [1984] and Tijms [1986]). More important is the result that the actual values of the β-service level for the various demand distributions differ only slightly. Hence we conclude that for the case of $0<c_D^2\leq 1$ the service level of the approximate (m,M)-rule is fairly insensitive to more than the first two moments of the underlying demand distribution. This implies that for the case of positive setup times the two-moment approximations, using mixtures of Erlangian distributions, yield practically useful results.

In table 6.5.3 we consider again the (m,M)-rules computed in table 6.5.1 and compare V_{app} and V_{act}, respectively the approximate and actual values of the average on-hand inventory. The approximate values of the average on-hand inventory are computed from approximation 6.4.1, where for the case of deterministic demand we used the exact undershoot distribution. The results show an excellent performance of the approximation 6.4.1, including for deterministic demand and for the case of $c_D^2>0$ and M=m.

Finally we turn our attention to the solution of problem Pb associated with the case of M-m>0. Having established the accuracy of the approximations for the service measures and the average on-hand inventory, we can use these approximations to find an approximately average-cost optimal (m,M)-rule that satisfies a given service level constraint. In section 6.4 we argued that the optimal value Δ^* of the difference M-m converges to $\tilde{\Delta}^*$ as the service level approaches 1. Here the quantity $\tilde{\Delta}^*$ is computed as a positive solution of the equation $z(\Delta)=0$, where $z(\Delta)$ is defined in section 6.4. However, in the computation of the approximately optimal (m,M)-rules we use the distribution functions $\hat{F}_\xi(x)$ and $\hat{F}_\eta(x)$ instead of the true $F_\xi(x)$ and $F_\eta(x)$. Therefore we can only hope that the approximate Δ^* gets close to $\tilde{\Delta}^*$.

Table 6.5.3. The approximate average inventories and their actual values.

			$c_D^2=0$			$c_D^2=1/3$		
π_2	β	K	m	V_{app}	V_{act}	m	V_{app}	V_{act}
1.25	0.95	0	9.55	6.29	6.31(3)	11.91	8.03	8.03(6)
2	0.95	0	5.96	4.16	4.15(1)	6.81	4.84	4.83(1)
5	0.95	0	5.38	3.95	3.93(1)	5.86	4.39	4.37(1)
1.25	0.99	0	13.31	9.99	10.00(5)	17.09	13.11	13.10(8)
2	0.99	0	7.96	6.13	6.12(2)	9.22	7.22	7.22(2)
5	0.99	0	7.30	5.84	5.84(1)	8.04	6.54	6.54(1)
1.25	0.95	25	8.17	6.74	6.73(5)	10.25	8.35	8.37(7)
2	0.95	25	4.66	5.56	5.54(2)	4.97	5.98	5.98(2)
5	0.95	25	3.65	5.77	5.75(1)	3.87	6.07	6.07(1)
1.25	0.99	25	11.94	10.45	10.49(5)	15.42	13.44	13.44(8)
2	0.99	25	6.76	7.64	7.64(2)	7.53	8.51	8.51(2)
5	0.99	25	5.68	7.77	7.76(1)	6.23	8.40	8.38(2)
			$c_D^2=2/3$			$c_D^2=2$		
π_2	β	K	m	V_{app}	V_{act}	m	V_{app}	V_{act}
1.25	0.95	0	14.31	9.80	9.82(9)	24.49	17.50	17.66(20)
2	0.95	0	7.72	5.59	5.60(3)	11.25	8.51	8.56(4)
5	0.95	0	6.43	4.92	4.91(1)	7.87	6.21	6.27(2)
1.25	0.99	0	20.89	16.26	16.21(12)	37.47	30.21	30.37(31)
2	0.99	0	11.70	8.54	8.53(2)	17.25	14.44	14.42(6)
5	0.99	0	9.04	7.51	7.50(1)	12.45	10.75	10.76(2)
1.25	0.95	25	12.58	10.07	10.11(12)	22.79	17.74	17.81(23)
2	0.95	25	5.73	6.60	6.59(2)	9.24	9.50	9.56(4)
5	0.95	25	4.28	6.46	6.47(2)	6.04	8.13	8.16(2)
1.25	0.99	25	19.16	16.53	16.58(11)	35.78	30.45	30.57(27)
2	0.99	25	8.86	9.70	9.69(2)	15.29	15.48	15.48(6)
5	0.99	25	7.08	9.24	9.25(2)	10.84	12.89	12.92(2)

In table 6.5.4 we display the approximately optimal values Δ^*. Also, we computed the quantity $\tilde{\Delta}^*$. We assume that excess demand is backlogged. In all examples $\lambda=E[D]=1$, $h=1$ and $K=25$. As before we consider a deterministic setup time T, which is varied as 0.5, 1, 2 and 4. For the values 0.99, 0.999 and 0.9999 of the required β-service level we computed the approximate solution to problem Pb by using approximation 6.4.1 for the expected holding cost per cycle. We consider Erlang-2 demand ($c_D^2=0.5$) and hyperexponential demand of order two with balanced means ($c_D^2=2$).

Table 6.5.4. Convergence of Δ^* to $\tilde{\Delta}^*$.

		$c_D^2=0.5$				$c_D^2=2$			
π_2	T	.99	.999	.9999	$\tilde{\Delta}^*$	.99	.999	.9999	$\tilde{\Delta}^*$
2	0.5	5.22	5.18	5.17	5.17	6.48	6.48	6.47	6.51
5	0.5	6.10	6.03	5.99	5.82	7.11	7.10	7.09	7.08
2	1	4.76	4.67	4.64	4.60	6.02	5.99	5.98	5.99
5	1	5.76	5.67	5.63	5.26	6.71	6.68	6.66	6.50
2	2	3.83	3.69	3.61	3.47	4.88	4.82	4.81	4.90
5	2	4.97	4.83	4.77	4.14	5.59	5.46	5.39	5.28
2	4	1.88	1.65	1.53	1.16	2.88	2.70	2.62	2.55
5	4	3.14	2.94	2.83	1.89	3.78	3.57	3.45	2.78

From the results in table 6.5.4 we can draw a number of conclusions. Firstly, in most cases the value of Δ^* gets closer to $\tilde{\Delta}^*$ as β increases. This was argued in section 6.4 for the true $F_\xi(x)$ and $F_\eta(x)$. Since we use the approximated distributions $\hat{F}_\xi(x)$ and $\hat{F}_\eta(x)$ instead of $F_\xi(x)$ and $F_\eta(x)$, this is another illustration of the power of the two-moment approximations suggested in section 6.2. Convergence to $\tilde{\Delta}^*$ is not guaranteed as follows from the three cases with $c_D^2=2$, $\pi_2=2$ and T has either of the values 0.5, 1 and 2. However, we found that in all cases considered Δ^* converges to some constant as β increases. Secondly, the rate of convergence decreases as T increases. This was already argued heuristically in the previous section. Thirdly, Δ^* decreases as T increases; the same holds for $\tilde{\Delta}^*$.

Denote by g_{EOQ} and $\tilde{g}^*$ the average switching and holding costs of the (m,M)-rules that are obtained for the β-service level requirement when using (1.5.1) for M-m, respectively $\tilde{\Delta}^*$ for M-m. In table 6.5.5 we compare g_{EOQ} and $\tilde{g}^*$ with the minimal average costs g^* for the (m,M)-rules obtained in table 6.5.4 with β=0.99. The costs g_{EOQ}, $\tilde{g}^*$ and g^* are computed from approximation 6.4.1 and equation (6.4.10).

As in table 5.6.3 we observe that the (m,M)-rule with M-m given by formula (1.5.1) performs quite well in costs. Our numerical investigations indicate that the relative error $(g_{EOQ}-g^*)/g^*$ is usually below 5% (see also De Kok et al [1985]), while for the (M-m)-rule with M-m=$\tilde{\Delta}^*$ the relative deviation from the minimal costs is negligible and is typically below 1%. Therefore we recommend the use of $\tilde{\Delta}^*$.

Table 6.5.5. Comparison of g_{EOQ}, $\tilde{g}^*$ and g^*.

		$c_D^2=0.5$			$c_D^2=2$		
π_2	T	g_{EOQ}	$\tilde{g}^*$	g^*	g_{EOQ}	$\tilde{g}^*$	g^*
2	0.5	9.22	9.21	9.21	17.01	16.93	16.93
5	0.5	9.09	9.09	9.08	14.39	14.36	14.36
2	1	9.63	9.63	9.63	17.33	17.30	17.30
5	1	9.70	9.70	9.68	14.98	14.98	14.98
2	2	10.44	10.38	10.37	17.80	17.80	17.80
5	2	10.71	10.67	10.63	15.82	15.80	15.80
2	4	11.89	11.55	11.52	19.20	19.09	19.09
5	4	12.33	12.05	11.97	17.78	17.63	17.60

We observe that $\tilde{\Delta}^*$ and Δ^* decrease as the deterministic setup time increases. Recommending the use of $\tilde{\Delta}^*$, we must have $\tilde{\Delta}^* \geq \Delta_0$ with Δ_0 defined by (6.3.19). There exists some T, say T_0, such that $\tilde{\Delta}^* = \Delta_0$. If T increases further a positive solution to $z(\Delta)=0$ is less than Δ_0 or does not exist. However, the function $z(\Delta)$ is only meaningful when $\Delta \geq \Delta_0$. Thus we suggest the following procedure.

1. Find a positive solution to $z(\Delta)=0$.
2. If this solution $\tilde{\Delta}^*$ exists and $\tilde{\Delta}^* \geq \Delta_0$, then set M-m equal to $\tilde{\Delta}^*$ and solve for the switching level m using the service level constraint.
3. If $\tilde{\Delta}^* < \Delta_0$ or there is no positive solution to $z(\Delta)=0$, then solve for the switching level m for each of the cases M-m=0 and $M-m=\Delta_0$. Use the (m,M)-rule which yields the lowest costs.

The setup time T_0 for which $\tilde{\Delta}^* = \Delta_0$ can be determined numerically from the function $z(\Delta)$. Because of $c_\delta(T)$, the function $z(\Delta)$ is also a function of T. Since T is deterministic we have

$$c_\delta(T) = \frac{\pi_2}{\lambda E[D]} e^{\pi_2 \delta T}.$$

Then T_0 can be solved as a positive root of $w(T)=0$ with $w(T)$ defined by

$$w(T) := h\Delta_0(\Delta_0+E[U]+\lambda E[D]T) - \frac{h}{\delta}(\Delta_0+E[U]+\lambda E[D]T)$$
$$+ h(\Delta_0+E[U]+\lambda E[D]T)^2 e^{-\delta\Delta_0}(\pi_2 e^{\pi_2\delta T}/(\lambda E[D]) - e^{-\delta\Delta_0})^{-1}$$
$$- \frac{h}{2}[\Delta_0^2 - E[U^2] - (2E[U]\lambda E[D]+\lambda E[D^2])T - (\lambda E[D])^2T^2]$$
$$- K\lambda E[D](\pi_2-\lambda E[D])/\pi_2,$$

with $E[U]$ and $E[U^2]$ given by (6.4.3) and (6.4.4). Hence if $T>T_0$ then we need to carry out only step 3 of the above procedure.

Conclusions.

The approximations for the service measures show an excellent performance for the case of $c_D^2>0$ and $M-m\geq\Delta_0$. If $c_D^2=0$ or $M=m$ then the performance of the approximations is of an acceptable quality provided $\lambda E[D]/\pi_2\geq 0.5$.

For fixed $M-m$ the switching level m satisfying a given service level constraint is quite insensitive to more than the first two moments of the demand size distribution provided $0<c_D^2\leq 1$. A two-moment approach, based on fitting a mixture of Erlang-k and Erlang-(k-1) demand distributions with the same scale parameters to the first two moments of the demand size, leads to practically useful results.

The approximation for the average on-hand inventory based on approximation 6.4.1 shows an overall excellent performance.

Assuming linear holding costs and fixed setup costs, the (m,M)-rule with $M-m$ determined by the EOQ-formula (1.5.1) and with m determined by the service level constraint shows a good performance in costs. Provided a positive solution $\tilde{\Delta}^*$ to $z(\Delta)=0$ exists with $\tilde{\Delta}^*\geq\Delta_0$, an improvement in costs results from the use of the (m,M)-policy with $M-m$ equal to $\tilde{\Delta}^*$. Another reason to prefer the choice $\tilde{\Delta}^*$ for $M-m$ rather than the EOQ-formula (1.5.1) not involving T is the fact that $\tilde{\Delta}^*$ decreases as T increases. This qualitative behaviour was also found empirically for the optimal value of $M-m$.

7. A DAM PROBLEM WITH VARIABLE RELEASE RATE.

In this chapter we apply results obtained in previous chapters to a dam problem, in which the input process is a compound Poisson process and the content of the dam can be released at two different rates. For this model we want to determine a control rule that gives an appropriate balance between the two unfavourable phenomena of overflow and emptiness. Clearly these two phenomena are conflicting in the sense that preventing overflows may cause the dam to be empty too often, while preventing an empty dam may contribute to overflows. In most practical situations it is hard to specify costs for emptiness and overflow. Therefore we focus on commonly used service measures. The service measures to consider are the probability of emptiness, the fraction of input that is lost and the average number of upcrossings of a certain critical level per unit time. Such a critical level may be relevant in the situation in which the dam has a finite capacity and overflows are temporarily stored elsewhere.

Again we assume that the content is controlled by an (m,M)-rule, which is described below. For given values of m and M, we derive tractable expressions for the service measures. Assuming that M-m is predetermined we can calculate m such that a prespecified service level is satisfied.

7.1. The model.

The dam model can be described as follows. The dam has a (possibly infinite) capacity C. Inputs into the dam occur at time epochs that form a Poisson process with rate λ. The inputs are independent random variables having a common probability distribution function F with $F(0)=0$ and finite second moment. The inputs are independent of the arrival process. The release of the dam content can be controlled by using one out of two possible release rates σ_1 and σ_2 with $\sigma_1 \leq \sigma_2 \leq \infty$. Under release rate σ_i the content decreases linearly at rate σ_i between input epochs as long as the content is positive. We assume that the content is controlled by an (m,M)-rule with $0 \leq m \leq M \leq C$. Under this rule the release rate is switched from σ_2 to σ_1 as soon as the content decreases to m. The release rate is switched from σ_1 to σ_2 as soon as the content becomes larger than M. We assume that

$$\sigma_1 < \lambda E[A] < \sigma_2, \tag{7.1.1}$$

where the random variable A denotes the size of a single input. For the case of $\sigma_1=0$ we assume that F is non-arithmetic.

There exists an extensive literature on the dam problem without control. For more general release functions first passage time results are derived by Yeo [1975], Brockwell and Chung [1975] and Ali Khan [1977] amongst others. The stationary distribution of the content was studied by Moran [1969], Brockwell [1977] and Smith and Yeo [1981]. In the last paper numerically tractable results are obtained.

The dam problem with controllable release rate will in general not allow for an exact analysis that leads to tractable results useful for practical applications. Diffusion process approximations are studied in Faddy [1974], Zuckermann [1977] and Attia and Brockwell [1982] for the dam problem with $\sigma_1=0$ and m=0. In these papers attention is focussed on the minimization of costs. Under a particular cost structure their analysis leads to a simple rule for the determination of M. By their continuous nature diffusion process approximations cannot adequately deal with service measures whose values are intrinsically determined by a jump process. Our approach based on renewal and random walk theory enables us to deal with such service measures as the average number of overflows per unit time. For the special case of exponentially distributed input tractable results for both the average holding and switching costs and the service levels of an (m,M)-rule can be deduced from Tijms and Van der Duyn Schouten [1978].

This chapter is further organized as follows. In section 7.2 we discuss the service measures and introduce the basic functions in which the service measures can be expressed. Next in section 7.3 we use approximations for the backlog and lost-sales model derived in the chapters 1 and 2, respectively, to obtain approximations for some of these basic functions. In section 7.4 approximations are derived for the other basic functions. In section 7.5 we derive an expression for the average content of the dam. Section 7.6 concludes this chapter with the presentation of numerical results.

7.2. The service measures.

As stated in the introduction we focus on several widely used service measures. In this section we will use results from the theory of regenerative processes to express the service measures in a number of basic functions. For these functions asymptotic estimates can be derived. We shall consider the following service measures:

1. the long-run fraction of time that the dam is empty.
2. the long-run fraction of input that is lost by overflows.
3. the average number of upcrossings of a critical level U per unit time.

Here an upcrossing of the level U means that the content reaches or exceeds U from below.

Assuming that the difference M-m is predetermined, the goal is to find the switching level m such that a prespecified value is achieved by one of these service levels. Our analysis is as follows. We first derive approximate expressions for the various service levels under a *given* (m,M)-rule. Next these expressions enable us to determine the switching level m in order to achieve a prespecified value of the service level. We give a unified treatment of the finite and infinite model.

Now we fix an (m,M)-rule. Also, let the critical level U be such that $M \leq U \leq C$. For any $t \geq 0$, define

$X(t)$:= the content of the dam at time t.

$V(t)$:= the total amount of input during $(0,t]$.

$L(t)$:= the amount of input that is lost by overflows during $(0,t]$.

$T_E(t)$:= the amount of time that the dam is empty during $(0,t]$.

$N_U(t)$:= the number of upcrossings of the level U during $(0,t]$.

Note that $L(t)$ is identically equal to zero when the capacity $C=\infty$. Due to assumption (7.1.1) the process $\{X(t), t \geq 0\}$ is a regenerative stochastic process, so that we can relate the long-run behaviour of the above defined random variables to their behaviour during a regeneration cycle. We define the regeneration cycle as the time elapsed between two consecutive epochs at which the release rate is switched from σ_2 to σ_1. Unless stated otherwise we assume that at epoch 0 a cycle starts. Define

T := the next epoch at which the release rate is switched from σ_2 to σ_1.

V := the total amount of input during $(0,T]$.

L := the amount of input that is lost by overflows during $(0,T]$.

T_E := the amount of time the dam is empty during $(0,T]$.

N_U := the number of upcrossings of U during $(0,T]$.

By a standard result from the theory of regenerative processes (cf. Cohen [1976] and Ross [1970]), we have with probability 1

$$\lim_{t\to\infty} \frac{T_E(t)}{t} = \frac{E[T_E]}{E[T]} \tag{7.2.1}$$

$$\lim_{t\to\infty} \frac{L(T)}{V(T)} = \frac{E[L]}{E[V]} \tag{7.2.2}$$

$$\lim_{t\to\infty} \frac{N_U(t)}{t} = \frac{E[N_U]}{E[T]} \tag{7.2.3}$$

Note that the left hand sides of (7.2.1), (7.2.2) and (7.2.3) represent respectively the long-run fraction of time the dam is empty, the long-run fraction of input that is lost and the average number of upcrossings of U per unit time. Also, we have for the average input per unit time

$$\frac{E[V]}{E[T]} = \lambda E[A].$$

Hence it suffices to derive tractable expressions for $E[T]$, $E[T_E]$, $E[L]$ and $E[N_U]$. Therefore we introduce a number of basic functions. Assuming that at epoch 0 the content $X(0)=x$, $0\leq x\leq M$, and release rate σ_1 is used, define

$t_1(x)$:= the expected time until the first upcrossing of the level M by $\{X(t),\ t\geq 0\}$.

$t_E(x)$:= the expected amount of time the dam is empty until the first upcrossing of the level M by $\{X(t),\ t\geq 0\}$.

$p(x,u)$:= the probability that the content just prior to the arrival of the input causing the first overshoot of the level M plus this input is at least $u-M$, $u\geq 0$.

The definition for $p(x,u)$ is rather subtle in order to unify the analysis for both the infinite and the finite capacity model. In the infinite capacity case $p(x,u)$ is just the probability that the switching level M is overshot by at least u. Next we define some functions that describe the system under release rate σ_2. Assuming that at epoch 0 the content $X(0)=x$, $M\leq x\leq C$, and release rate σ_2 is used, define

$t_2(x)$ = the expected time until the content decreases to m.

$\phi(x)$ = the expected amount of input that is lost by overflow until the content decreases to m.

$n_U(x)$ = the expected number of upcrossings of U until the content decreases to m.

Using the basic functions it is easy to see that the following relations hold. In the case of $C<\infty$ we have

$$E[T] = t_1(m) + \int_0^{C-M} t_2(M+u)d_u(1-p(m,u))+t_2(C)p(m,C-M). \tag{7.2.4}$$

$$E[T_E] = t_E(m). \tag{7.2.5}$$

$$E[L] = \int_0^{C-M} \phi(M+u)d_u(1-p(m,u))+ \int_{C-M}^{\infty} [u+\phi(C)]d_u(1-p(m,u)). \tag{7.2.6}$$

$$E[N_U] = \int_0^{U-M} n_U(M+u)d_u(1-p(m,u))+[n_U(U)+1]p(m,U-M). \tag{7.2.7}$$

In the above integrals the integration interval is right-open. In the case of $C=\infty$ we have $E[L]=0$ and (7.2.4) changes into

$$E[T] = t_1(m) + \int_0^{\infty} t_2(M+u)d_u(1-p(m,u)). \tag{7.2.8}$$

Relations (7.2.5) and (7.2.7) hold in the infinite capacity model too. Actually, since $U\leq C$ it is easily seen that $n_U(x)$ does not depend on the capacity C of the system. This follows by making the following two observations. First, after each upcrossing of U the inventory has to decrease to U first before the next upcrossing can occur. Second, since customers arrive according to a Poisson process the epoch at which the inventory decreases to U can be considered as an epoch immediately after

an arrival. In section 7.3 we will give approximations for the functions $t_2(x)$, $\phi(x)$ and $n_U(x)$. These approximations result from a relation between the dam model and the production-inventory models with controllable production rate studied in the chapters 1 and 2. In section 7.4 approximations are obtained for $t_1(x)$, $t_E(x)$ and $p(x,u)$.

We end this section with the introduction of some quantities that will be needed in the sequel. As before let $s^*>0$ be the unique strictly positive solution to

$$(7.2.9) \qquad s - \frac{\lambda}{\sigma_1}(1-\tilde{F}(s)) = 0,$$

where $\tilde{F}$ denotes the Laplace-Stieltjes transform of F. Then we can define

$$(7.2.10) \qquad G(x) := \int_0^x e^{-s^*y} \frac{\lambda}{\sigma_1}(1-F(y))dy.$$

Becuase of (7.2.9) G is a proper distribution function. Next we assume that the equation in $t\geqq 0$

$$(7.2.11) \qquad \int_0^\infty e^{ty} \frac{\lambda}{\sigma_2}(1-F(y))dy = 1$$

has a solution $\delta>0$, which is necessarily unique. We noted that a necessary and sufficient condition for (7.2.11) to have a solution is

$$1-F(x) = O(e^{-\kappa x}) \quad \text{for some } \kappa > 0 \ (x\to\infty).$$

Also, we define the finite constant

$$\nu := \int_0^\infty ye^{\delta y} \frac{\lambda}{\sigma_2}(1-F(y))dy.$$

Finally, let a, b and ξ be defined by

$$a := \lambda E[A]/\sigma_2 - (\delta\nu\sigma_2)^{-1}(\sigma_2-\lambda E[A]),$$

$$b := \tfrac{1}{2}(\sigma_2-\lambda E[A])^{-1}\lambda E[A^2] - (\delta^2\nu\sigma_2)^{-1}(\sigma_2-\lambda E[A]),$$

$$\xi := a/b$$

and the functions $h_1(y)$ and $h_2(y)$ by

$$h_1(y) := \sigma_2 a(\sigma_2-\lambda E[A])^{-1}e^{-\xi y}+(\delta\nu)^{-1}e^{-\delta y}, \quad y \geq 0,$$

$$h_2(y) := be^{-\xi y}+(\sigma_2-\lambda E[A])(\delta^2\nu\sigma_2)^{-1}e^{-\delta y}, \quad y \geq 0.$$

7.3. Approximations for $t_2(x)$, $\phi(x)$ and $n_U(x)$.

In this section we will derive approximations for the basic functions that describe the behaviour of the system under release rate σ_2. In the chapters 1 and 2 we analyzed one-product production-inventory models in which customers arrive according to a Poisson process with rate λ. The demands of the customers are independent random variables having a common probability distribution function F with F(0)=0 and the demands are independent of the arrival process. The commodity is continuously produced at a rate $\sigma_2>0$. The system has an infinite storage capacity. Letting D denote the demand size per customer, it is assumed that $\sigma_2>\lambda E[D]$. We relate the basic functions $t_2(x)$, $\phi(x)$ and $n_U(x)$ to analogous functions that appear in these production inventory models. For the latter functions approximations have been derived in chapter 1 and 2.

Consider first the infinite capacity dam problem and the inventory problem with backlogging of excess demand. It is easily seen that the expected time $t_2(x)$ needed to reduce the content from x to m under release rate σ_2 equals the expected time until the inventory has increased by an amount of x-m under production rate σ_2. From equation (1.4.1) we have for the case of $C=\infty$

$$(7.3.1) \qquad t_2(x) = \frac{x-m}{\sigma_2-\lambda E[A]}, \quad x \geq m.$$

Next we consider the finite capacity dam problem and the inventory problem in which excess demand is lost. For the lost-sales model we define for all $M\geq 0$ and $0\leq x\leq M$,

$\ell_M(x)$:= the expected amount of demand lost until the inventory level increases to M, given that the initial inventory equals x.

It follows after some reflection that

$$(7.3.2) \qquad \phi(x) = \ell_{C-m}(C-x), \quad m \leq x \leq C.$$

The approximation 2.3.2 and the definition of $h_2(x)$ together imply

$$\ell_M(x) \cong h_2(x) - h_2(M).$$

Hence, using (7.3.2) we obtain

$$\phi(x) \cong h_2(C-x) - h_2(C-m), \quad m \leq x \leq C. \tag{7.3.3}$$

Next we apply the principle of conservation of flow to obtain a relation between $t_2(x)$ and $\phi(x)$ for the finite capacity model. The conservation of flow states that for all $t \geq 0$ the content $X(t)$ equals the sum of the initial content $X(0)$ and the amount of input that does not overflow in $(0,t]$ minus the amount released in $(0,t]$. Using this principle and applying Wald's equation we obtain

$$m = x + \lambda E[A]t_2(x) - \phi(x) - \sigma_2 t_2(x), \quad m \leq x \leq C.$$

Rearranging terms and using (7.3.3) yields

$$t_2(x) \cong \frac{x-m-[h_2(C-x)-h_2(C-m)]}{\sigma_2-E[A]}, \quad m \leq x \leq C. \tag{7.3.4}$$

Finally we derive an approximation for $n_U(x)$. We recall that $n_U(x)$ is the same for both the finite and infinite capacity model provided $U \leq C \leq \infty$. Let us consider the infinite capacity model. Comparing the infinite capacity model with the backlog model we observe that $n_U(x)$ equals the expected number of stockout occurrences until the inventory reaches the level $U-m$, given that the initial inventory equals $U-x$, $m \leq x \leq U$. For the latter function we can derive an approximation by applying the same arguments as used to derive (1.2.11). Combination of approximation 1.4.1 and the definition of $h_1(x)$ then yields

$$n_U(x) \cong \begin{cases} h_1(U-x)-h_1(U-m), & m \leq x \leq U \\ h_1(0)-h_1(U-m), & U < x \leq C \end{cases} \tag{7.3.5}$$

In the next section we derive approximations for the basic functions that describe the system under release rate σ_1.

7.4. Approximations for $t_1(x)$, $t_E(x)$ and $p(x,u)$.

In this section the functions $t_1(x)$, $t_E(x)$ and $p(x,u)$ related to the use of release rate σ_1 are approximated. Since the content of the dam cannot become negative the dam problem under production rate σ_1 corresponds to a production inventory problem with finite capacity. For the case of $\sigma_1 \leqq 0$ the infinite capacity model is in fact a finite capacity model, since the inventory level cannot exceed the highest switching level. Therefore expressions for $t_1(x)$, $t_E(x)$ and $p(x,u)$ can be deduced from results in chapter 1. Below we will give these results without further explanation. However, for the case of $\sigma_1 > 0$ the analysis in chapter 1 essentially uses an infinite storage capacity. Thus, in developing approximations for $t_1(x)$, $t_E(x)$ and $p(x,u)$ for the case of $\sigma_1 > 0$ we cannot simply use existing approximations from chapter 1.

Assuming that at epoch 0 the content equals x and release rate $\sigma_1 > 0$ is used, we define the following random variables,

$T_1(x)$:= the time until the content upcrosses M.

$T_E(x)$:= the amount of time the dam is empty during $(0,T_1(x)]$.

$U(x)$:= the amount by which the sum of the content just prior to time $T_1(x)$ and the input at time $T_1(x)$ exceeds M.

$N(x)$:= the number of inputs during $(0,T_1(x)]$.

τ_n := the time elapsed between the $(n-1)^{th}$ and n^{th} input, $n \geqq 1$ (assume that the 0^{th} input occurs at epoch 0).

A_n := the size of the n^{th} input.

We first derive a relation between $t_1(x)=E[T_1(x)]$, $E[U(x)]$ and $t_E(x)=E[T_E(x)]$. At time $T_1(x)$ the $\{N(x)\}^{th}$ input is stored into the dam and the sum of the content just prior to time $T_1(x)$ and $A_{N(x)}$ equals $M+U(x)$. During $(0,T_1(x)]$ an amount of $\sum_{i=1}^{N(x)} A_i$ is added to the initial content x, while the total amount released during $(0,T_1(x)]$ equals $\sigma_1(T_1(x)-T_E(x))$. Thus, by the principle of conservation of flow,

$$(7.4.1) \qquad M+U(x) = x + \sum_{i=1}^{N(x)} A_i - \sigma_1(T_1(x)-T_E(x)).$$

By the definition of $T_1(x)$,

$$(7.4.2) \qquad T_1(x) = \sum_{i=1}^{N(x)} \tau_i.$$

It is easy to see that the event $\{N(x)=n\}$ is independent of $\{T_{n+1}, A_{n+1}, \ldots\}$. Applying Wald's equation, we obtain

$$(7.4.3) \qquad E[\sum_{i=1}^{N(x)} A_i] = E[A]E[N(x)]$$

and

$$(7.4.4) \qquad E[\sum_{i=1}^{N(x)} \tau_i] = (1/\lambda)E[N(x)].$$

Combining (7.4.1)-(7.4.4) we can express $t_1(x)$ in terms of $t_E(x)$ and $E[U(x)]$,

$$(7.4.5) \qquad t_1(x) = \frac{M-x+E[U(x)]-\sigma_1 t_E(x)}{\lambda E[A]-\sigma_1}.$$

Note that

$$E[U(x)] = \int_0^\infty p(x,u)du.$$

Hence it suffices to find approximations for $t_E(x)$ and $p(x,u)$. We first derive a renewal equation related to $t_E(x)$. From an approximate solution of this equation we find an approximation for $t_E(x)$.

Conditioning on the possible events in $(0, \frac{\Delta x}{\sigma_1})$ with Δx small we derive the following relation

$$t_E(x) = (1-\frac{\lambda\Delta x}{\sigma_1})t_E(x-\Delta x) + \frac{\lambda\Delta x}{\sigma_1}[\int_0^{M-x} t_E(x+y)dF(y)]+o(\Delta x),$$

$$0 < x \leq M.$$

Rearranging terms and letting Δx approach zero we find an integro-differential equation for $t_E(x)$

$$(7.4.6) \qquad t_E'(x) = -\frac{\lambda}{\sigma_1} t_E(x) + \frac{\lambda}{\sigma_1} \int_0^{M-x} t_E(x+y)dF(y), \quad 0 < x \leq M.$$

A boundary condition is obtained for $t_E(0)$ by using the fact that customers arrive according to a Poisson process with rate λ. Hence

$$t_E(0) = \frac{1}{\lambda} + \int_0^M t_E(y)dF(y).$$

Using a technique given in Feller [1971], we rewrite (7.4.6) into a renewal equation. For this purpose we introduce an auxiliary function w, which is defined by

$$(7.4.7) \qquad w(x) := t_E(M-x), \quad 0 \leq x \leq M.$$

Then we have the following equations for w

$$(7.4.8) \qquad w'(x) = \frac{\lambda}{\sigma_1} w(x) - \frac{\lambda}{\sigma_1} \int_0^x w(x-y)dF(y), \quad 0 \leq x < M.$$

$$(7.4.9) \qquad w(M) = \frac{1}{\lambda} + \int_0^M w(M-y)dF(y).$$

Using the differentiation rule for an integral and using partial integration with $F(0)=0$, it is easy to verify that

$$\frac{d}{dx} \int_0^x w(x-y)(1-F(y))dy = w(x) - \int_0^x w(x-y)dF(y).$$

Hence (7.4.8) can be rewritten as

$$(7.4.10) \qquad w(x) = w(0) + \frac{\lambda}{\sigma_1} \int_0^x w(x-y)(1-F(y))dy, \quad 0 \leq x < M.$$

Multiplying (7.4.10) by e^{-s^*x} and using (7.2.10) we obtain the renewal equation

$$w(x)e^{-s^*x} = w(0)e^{-s^*x} + \int_0^x w(x-y)e^{-s^*(x-y)}dG(y), \quad 0 \leq x < M.$$

This equation has the unique solution

$$w(x)e^{-s^*x} = \int_0^x w(0)e^{-s^*(x-y)}dR(y), \quad 0 \leq x < M,$$

where the renewal function R(x) is defined by

$$(7.4.11) \qquad R(x) = \sum_{n=0}^{\infty} G^{n*}(x),$$

with $G^{n*}(x)$ denoting the n-fold convolution of G with itself. Thus we find

$$(7.4.12) \qquad w(x) = w(0) \int_0^x e^{s^* y} dR(y), \quad 0 \leq x < M.$$

Noting that w(x) is continuous at x=M and substituting (7.4.12) into (7.4.9), we can solve for w(0). After some algebra we obtain

$$w(0) = \{\lambda \int_0^M e^{s^* y}(1-F(M-y))dR(y)\}^{-1}.$$

The expression between brackets on the right-hand side of this equation can be simplified further. Using the definition of G(x) given by (7.2.10) we find

$$\lambda \int_0^M e^{s^* y}(1-F(M-y))dR(y) = \sigma_1 e^{s^* M} \int_0^M G'(M-y)dR(y).$$

Here G'(x) denotes the first derivative of G. Since R(x) is the renewal function associated with G(x) we have

$$\int_0^x G(x-y)dR(y) = R(x)-1.$$

Taking derivatives on either sides we obtain

$$G(0)\, R'(x) + \int_0^x G'(x-y)dR(y) = R'(x).$$

Using G(0)=0 the above arguments lead to the following equality

$$\lambda \int_0^M e^{s^* y}(1-F(M-y))dR(y) = \sigma_1 e^{s^* M} R'(M)$$

and thus we find

$$(7.4.13) \qquad w(0) = \{\sigma_1 e^{s^* M} R'(M)\}^{-1}.$$

Together, (7.4.7), (7.4.12) and (7.4.13) determine the function $t_E(x)$.

However, we cannot obtain useful results from (7.4.12) and (7.4.13), since a tractable expression for R(x) is in general not available. Therefore we derive an approximation for R(x).

Letting B be a random variable having distribution function G, a well-known result from renewal theory (cf. Ross [1970]) states that

$$(7.4.14) \qquad \lim_{x\to\infty} [R(x) - \{\frac{x}{E[B]} + \frac{E[B^2]}{2(E[B])^2}\}] = 0.$$

It follows from (7.2.10) that

$$(7.4.15) \qquad E[B] = (1+\lambda\tilde{F}'(s^*)/\sigma_1)/s^* \qquad E[B^2] = (-\lambda\tilde{F}''(s^*)/\sigma_1+2E[B])/s^*.$$

Hence, for x sufficiently large, (7.4.14) provides an approximation for R(x). Next we describe the behaviour of R(x) near the origin. We know that R(0)=1. We want our approximation $\hat{R}(x)$ of R(x) to be such that

$$(7.4.16) \qquad \hat{R}(0) = 1, \quad \hat{R}'(0) = R'(0).$$

Since G(0)=0, $G'(0)=\lambda/\sigma_1$ and G is (right-)continuously differentiable it follows that

$$(7.4.17) \qquad R'(0) = G'(0) = \lambda/\sigma_1.$$

There are several ways to choose $\hat{R}(x)$ such that $\hat{R}(x)$ satisfies (7.4.14) and (7.4.16). Our choice is based on the explicit expression for R(x) which can be found in case the input distribution function F belongs to the class of K_2-distribution functions. Recall that the probability distribution function $\tilde{F}$ is a K_2-distribution if its Laplace-Stieltjes transform F is of the following form,

$$(7.4.18) \qquad \tilde{F}(s) = \frac{1+(a_1-E[A])s}{1+a_1s+a_2s^2}, \quad s \geq 0$$

for some constants a_1 and a_2 with $a_2>0$. In view of (7.4.11) and (7.2.10) the Laplace-Stieltjes transform $\tilde{R}$ of R is given by

$$(7.4.19) \qquad \tilde{R}(s) = \frac{1}{1-\tilde{G}(s)} = \frac{\sigma_1(s+s^*)}{\sigma_1(s+s^*)-\lambda(1-\tilde{F}(s+s^*))}.$$

Combining (7.4.18) and (7.4.19) we find for the case of F belonging to the class of K_2-distributions

$$(7.4.20)\qquad \tilde{R}(s) = \sigma_1[1+a_1(s+s^*)+a_2(s+s^*)^2] \times$$

$$\times \{\sigma_1[1+a_1(s+s^*)+a_2(s+s^*)^2]-\lambda[a_2(s+s^*)+E[A]]\}^{-1}.$$

Using the definition of s^* we can rewrite (7.4.20) into

$$(7.4.21)\qquad \tilde{R}(s) = \lambda/\sigma_1[a_2(s+s^*)+E[A]]s^{-1}[a_2s+a_1+2a_2s^*-\lambda a_2/\sigma_1]^{-1}+1.$$

Noting that the denominator in (7.4.21) has two zeroes, we can invert $\tilde{R}(s)$ and after some algebra we find

$$(7.4.22)\qquad R(x) = 1 + c_1x + c_2(1-e^{-\gamma x}), \quad x \geq 0,$$

with

$$\gamma = [(\lambda a_2/\sigma_1-a_1)^2+4a_2(\lambda E[A]/\sigma_1-1)]^{\frac{1}{2}}/a_2,$$

$$c_1 = \lambda(s^*+E[A]/a_2)(\sigma_1\gamma)^{-1}, \quad c_2 = \lambda(\sigma_1\gamma)^{-1}-c_1/\gamma.$$

We take the right side of (7.4.22) as an approximate formula for R(x). The constants c_1, c_2 and γ follow by using (7.4.14) and (7.4.17). Thus, we suggest the approximation

$$(7.4.23)\qquad \hat{R}(x) = 1 + c_1x + c_2(1-e^{-\gamma x}), \quad x \geq 0,$$

with

$$(7.4.24)\qquad \gamma = 2(\lambda/\sigma_1-1/E[B])(c_B^2-1)^{-1},$$

$$(7.4.25)\qquad c_1 = 1/E[B], \quad c_2 = \tfrac{1}{2}(c_B^2-1),$$

where $c_B=\sigma(B)/E[B]$ denotes the coefficient of variation of B and E[B] and $E[B^2]$ are computed from (7.4.15) with the <u>true</u> F. Although this approximation ie exact in case (7.4.18) holds, some care should be taken in applying this

approximation. The approximation is only consistent with (7.4.14) if the constant γ is positive. A counter-example can be given showing that γ may be negative. However, for many practical input distributions we have that $\gamma>0$. The constant γ is not only positive for K_2-input, but it can also be shown that $\gamma>0$ both for deterministic input and for gamma input. In view of extensive numerical experiments we conjecture that $\gamma>0$ also holds in case the input distribution F is a mixture of Erlang-k and Erlang-(k-1) with the same scale parameters. A sufficient condition for $\gamma>0$ to hold is that G is NBUE (NWUE) and $E[B]\neq\lambda/\sigma_1$. In general one should numerically verify whether $\gamma>0$. For input distributions with $\gamma<0$ our numerical experiments reveal that good approximations for the switch-over levels are obtained when we replace γ by 0 in (7.4.23). By combining (7.4.7), (7.4.12), (7.4.13), (7.4.23), (7.4.24) and (7.4.25) an approximation for $t_E(x)$ is found.

For the case of $\sigma_1>0$ we have

$$t_E(x) \cong \frac{\int_0^{M-x} e^{s^* y} d\hat{R}(y)}{\sigma_1 e^{s^* M} \hat{R}'(M)}$$

with $\hat{R}(x)$ given by (7.4.23)-(7.4.25).
For the case of $\sigma_1<0$ we clearly have

$$t_E(x) = 0,$$

while for the case of $\sigma_1=0$

$$t_E(x) = \begin{cases} 1/\lambda & x = 0 \\ 0 & x > 0 \end{cases}$$

This leaves us with the problem to give tractable expressions for p(m,u). We use results obtained in chapter 1. In order to justify the application of these results we rephrase condition 1.3.1 in terms of the present dam model. We define

$$\zeta := \min\{n \mid \sum_{i=1}^{n} (A_i - \sigma_1 \tau_i) > 0\}.$$

$$Z_1 := \sum_{i=1}^{\zeta} (A_i - \sigma_1 \tau_i).$$

Hence $Z_1 = A_1 - \sigma_1 \tau_1$ if $\sigma_1 \leqq 0$. For the case of $\sigma_1 > 0$ the distribution of Z_1 is given by (1.3.20) with $\pi_1 = \sigma_1$. Then the (m,M)-rule must satisfy the following condition.

Condition 7.4.1.

For the case of $\sigma_1 > 0$ and M=m,

$$M \geqq \begin{cases} E[Z_1] & \text{when } c_{Z_1}^2 \leqq 1 \\ \frac{3}{2} c_{Z_1}^2 E[Z] & \text{when } c_{Z_1}^2 > 1 \end{cases}$$

For the case of M-m>0,

$$M-m \geqq \begin{cases} E[Z_1] & \text{when } c_{Z_1}^2 \leqq 1 \\ \frac{3}{2} c_{Z_1}^2 E[Z_1] & \text{when } c_{Z_1}^2 > 1 \end{cases}.$$

Here $c_Z^2 = \mathrm{var}(Z_1)/(E[Z_1])^2$.
This condition differs from condition 1.3.1 in that M must be sufficiently large for the case of M=m and $\sigma_1 > 0$. This difference is illuminated below.

We noted before that for the case of $\sigma_1 \leqq 0$ we can apply the results obtained in chapter 1 for the basic functions associated with the low production rate. From equation (1.3.34) we obtain for any $u \geqq 0$

$$p(M,u) = \begin{cases} 0 & \text{when } \sigma_1 < 0 \\ 1-F(u) & \text{when } \sigma_1 = 0 \end{cases}$$

From approximation 1.3.2 we obtain for the case of $\sigma_1 \leqq 0$ and M-m satisfying condition 7.4.1

$$p(m,u) \cong \frac{\lambda}{\lambda E[A] - \sigma_1} \int_u^{\infty} (1-F(y))dy, \quad u \geqq 0.$$

In case $\sigma_1>0$ we distinguish between the cases $c_A^2\leq 1$ and $c_A^2>1$ with $c_A=\sigma(A)/E[A]$. If $c_A^2\leq 1$ it is intuitively clear that the probability of a zero content just prior to the overshoot of M is negligible provided M is sufficiently large. In view of this observation we neglect the influence of the boundary at zero. Putting this in terms of the corresponding production-inventory model this is equivalent to neglecting the finite capacity if the difference between the capacity and the lowest switching level is sufficiently large. Then we use (1.3.20) and approximation 1.3.2, which give expressions for the distribution of the undershoot of the lowest switching level. Provided that the (m,M)-rule satisfies condition 7.4.1, we obtain for the case of $c_A^2\leq 1$,

$$(7.4.26)\qquad p(M,u) \cong \frac{\lambda}{\sigma_1}\int_u^\infty e^{-s^*(y-u)}(1-F(y))dy, \quad u \geq 0.$$

$$(7.4.27)\qquad p(m,u) \cong \frac{\lambda}{\lambda E[A]-\sigma_1}\int_u^\infty (1-e^{-s^*(y-u)})(1-F(y))dy, \quad u \geq 0.$$

Numerical experiments reveal that these approximations have an excellent performance. However if $c_A^2>1$ the same numerical experiments show that the performance of approximations (7.4.26) and (7.4.27) deteriorates if c_A^2 increases. If $c_A^2>1$ we typically have a large number of small inputs and a small number of very large inputs. The effect of the small inputs is on the average a net decrease of the content, while the large inputs may typically cause an overshoot. Roughly stated, in case $c_A^2>1$ we cannot neglect the possibility that an overshoot is caused by an input occurring when the content is close to zero. Thus in our analysis of p(x,u) we have to incorporate the boundary behaviour near x=0. To do this we shall proceed along the same lines as for the approximation of $t_E(x)$.

By conditioning on the possible events in $(0, \frac{\Delta x}{\sigma_1})$ and letting $\Delta x\to 0$ we find

$$\frac{\partial}{\partial x}p(x,u) = \frac{-\lambda}{\sigma_1}p(x,u) + \frac{\lambda}{\sigma_1}\int_0^{M-x} p(x+y,u)dF(y) +$$

$$+ \frac{\lambda}{\sigma_1}(1-F(M-x+u)), \quad 0 < x \leq M.$$

The boundary condition at x=0 is given by

$$(7.4.28)\qquad p(0,u) = 1-F(M+u) + \int_0^M p(y,u)dF(y).$$

We define another auxiliary function

$$(7.4.29) \qquad q(x,u) = p(M-x,u), \quad 0 \leq x \leq M.$$

Applying the same arguments as used to derive (7.4.12) we find

$$(7.4.30) \qquad q(x,u) = q(0,u) \int_0^x e^{s^* y} dR(y) - \int_0^x e^{s^* y} \int_0^{x-y} \frac{\lambda}{\sigma_1} (1-F(z+u)) dz dR(y) \qquad 0 \leq x \leq M.$$

Using (7.4.28)-(7.4.30) we obtain after considerable algebra

$$q(0,u) = \int_0^M e^{-s^*(M-y)} \frac{\lambda}{\sigma_1} (1-F(M-y+u)) dR(y) . \{R'(M)\}^{-1}.$$

Using the approximation $\hat{R}$ for R we suggest the following approximation for p(x,u) for the case of $\sigma_1 > 0$ and $c_A^2 > 1$,

$$(7.4.31) \qquad p(x,u) \cong \hat{q}(0,u) \int_0^{M-x} e^{s^* y} d\hat{R}(y) - \int_0^{M-x} e^{s^* y} \int_0^{M-x-y} \frac{\lambda}{\sigma_1} (1-F(z+u)) dz d\hat{R}(y), \quad 0 \leq x \leq M,$$

with

$$\hat{q}(0,u) = \int_0^M e^{-s^*(M-y)} \frac{\lambda}{\sigma_1} (1-F(M-y+u)) d\hat{R}(y) . \{\hat{R}'(M)\}^{-1}$$

and $\hat{R}(x)$ given by (7.4.23)-(7.4.25).

The approximation (7.4.31) for p(x,u) is less tractable than the other approximations given above. However, in case F is a hyper-exponential distribution function,

$$(7.4.32) \qquad F(x) = 1 - pe^{-\mu_1 x} - (1-p)e^{-\mu_2 x},$$

then the computations simplify considerably and yield a lengthy but tractable expression for p(x,u),

$$(7.4.33) \qquad p(x,u) = r_1(x)e^{-\mu_1 u} + (1-r_1(x))e^{-\mu_2 u},$$

where

$$r_1(x) = q_1(M)\{q_1(M)+q_2(M)\}^{-1}k_1(M-x) - \frac{\lambda}{\sigma_1}k_2(M-x)$$

with

$$q_i(M) = p_i\Big[\sum_{j=1}^{2}\frac{d_j}{\mu_i+\eta_j}(e^{\eta_j M}-e^{-\mu_i M}) + e^{-\mu_i M}\Big],$$

$$k_1(z) = \sum_{j=1}^{2}\frac{d_j}{\eta_j}(e^{\eta_j z}-1) + 1,$$

$$k_2(z) = \frac{p}{\mu_1}(1-e^{-\mu_1 z})+p\sum_{j=1}^{2}\Big[\frac{d_j}{\mu_1\eta_j}(e^{\eta_j z}-1)-\frac{d_j}{\mu_1(\eta_j+\mu_1)}(e^{\eta_j z}-e^{-\mu_1 z})\Big],$$

$p_1=p$, $p_2=1-p$, $d_1=1/E[B]$, $d_2=\lambda/\sigma_1-1/E[B]$, $\eta_1=s^*$ and $\eta_2=s^*-\gamma$.

The equality in (7.4.33) holds, since $\hat{R}(x)$ is exact for a hyperexponential distribution. Note $E[B]=(\lambda/\sigma_1)\sum_{i=1}^{2} p_i/(\mu_i+s^*)^2$ in case (7.4.32) holds.

In section 7.6 we present numerical results, showing the accuracy of the approximations to the levels of the service measures introduced in section 7.2.

7.5. The average content of the dam.

Assuming holding costs are incurred at a rate being linear in the stock on hand, the average holding cost per unit time follows by deriving an expression for the average content of the dam. Using results obtained in chapter 5 we derive approximate expressions for the average content.

Let us assume that a holding cost at rate x is incurred when the content level equals x. We define the random variable Z as

Z := the holding cost incurred during (0,T].

Then it follows from the theory of regenerative processes that

$$\text{the long-run average content of the dam} = \frac{E[Z]}{E[T]}.$$

To find an expression for E[Z] we define the functions $z_1(x)$ and $z_2(x)$ by

$z_1(x)$:= the expected holding costs incurred until the first overshoot of level M, when X(0)=x, $0 \leq x \leq M$, and the release rate σ_1 is used.

$z_2(x)$:= the expected holding costs incurred until the first epoch at which the content decreases to m, when X(0)=x, $x \geq m$, and the release rate σ_2 is used.

Then it is easily seen that

$$E[Z] = \begin{cases} z_1(m) + \int_0^{\infty} z_2(M+u)\, d_u(1-p(m,u)) & \text{when } C = \infty \\ z_1(m) + \int_0^{C-M} z_2(M+u)\, d_u(1-p(m,u)) + z_2(C)p(m,C-M) & \text{when } C < \infty. \end{cases} \tag{7.5.1}$$

To obtain expressions for the functions $z_1(x)$ and $z_2(x)$ we consider again the production-inventory model studied in chapter 1. However, we now assume that the storage capacity $\hat{C}$ is finite. We assume that a holding cost at rate x is incurred when the inventory level equals $x \geq 0$. Then we define for the backlog model

$k_1(x;\hat{C})$:= the expected holding costs incurred until the first epoch at which the inventory decreases below 0, given that at epoch 0 the inventory equals x, $0 \leq x \leq \hat{C}$, and production rate σ_1 is *always* used.

$k_2(x;M)$:= the expected holding cost incurred until the first epoch at which the inventory reaches the level $M \leq \hat{C}$, given that at epoch 0 the inventory equals x, $x \leq M$, and production rate σ_2 is *always* used.

We note that $k_2(x;M)$ does not depend on $\hat{C}$. Moreover, an approximation for $k_2(x;M)$ is given by equation (5.4.7) together with approximation 1.4.3, where we put π_2 equal to σ_2.

Some reflection will reveal that

$$z_1(x) = Mt_1(x) - k_1(M-x;M). \tag{7.5.2}$$

Using the definition of $k_2(x;M)$ and the arguments used to derive (5.4.7) we find for all $m \leq x \leq C$

$$(7.5.3) \qquad z_2(x) = \begin{cases} mt_2(x) + \frac{(x-m)^2}{2(\sigma_2-\lambda E[A])} + \frac{\lambda E[A^2]}{2(\sigma_2-\lambda E[A])^2}(x-m) & \text{when } C = \infty \\ Ct_2(x) - k_2(C-x;C-m) & \text{when } C < \infty \end{cases}$$

Because of (7.5.2) it remains to find an expression for $k_1(M-m;M)$. To do so let us assume that at epoch 0 the inventory equals $x \leq M$ and production rate σ_1 is always used. As stated before the finite capacity model can be treated as an infinite capacity model if the production rate $\sigma_1 \leq 0$. Then it follows from (5.2.6) and (5.3.26) that for the case of $\sigma_1 \leq 0$ and M-m satisfying condition 7.4.1

$$(7.5.4) \qquad k_1(M-m;M) \cong \frac{1}{\lambda E[A]-\sigma_1}\left[\frac{(M-m)^2}{2} - \frac{\lambda E[A^3]}{6(\lambda E[A]-\sigma_1)} + \frac{\lambda E[A^2]}{2(\lambda E[A]-\sigma_1)} \times \right.$$
$$\left. \times \left(M-m+ \frac{\lambda E[A^2]}{2(\lambda E[A]-\sigma_1)}\right)\right].$$

Clearly, $k_1(0;M)=0$ for the case of $\sigma_1 \leq 0$.

Let us consider the case of $\sigma_1 > 0$. Applying the renewal-theoretic arguments that were used to derive expressions for the auxiliary functions w(x) and q(x,u) we find after some straightforward algebra,

$$(7.5.5) \qquad k_1(x;M) = k_1(0;M) \int_0^x e^{s^* y} dR(y) - \int_0^x \frac{(x-y)^2}{2\sigma_1} e^{s^* y} dR(y), \quad 0 \leq x \leq M,$$

with

$$(7.5.6) \qquad k_1(0;M) = (\sigma_1 R'(M))^{-1} \int_0^M (M-y) e^{-s^*(M-y)} dR(y).$$

Substituting $\hat{R}(x)$ for R(x) into (7.5.5) and (7.5.6), we obtain a tractable approximation for $k_1(x;M)$, which is exact if the input distribution function F is a K_2-distribution. However, numerical investigations for deterministic input revealed that the accuracy deteriorates dramatically as σ_1 gets smaller for fixed m and M. Therefore we follow now another approach.

Equations (7.5.5) and (7.5.6) hold for all $0 \leq x \leq M$ and for all $M \geq 0$. Letting $M \to \infty$ we obtain for all $x \geq 0$

$$(7.5.7)\qquad k_1(x;\infty) = k_1(0;\infty)\int_0^x e^{s^*y}dR(y) - \int_0^x \frac{(x-y)^2}{2\sigma_1} e^{s^*y}dR(y).$$

Now we observe that the case of $M=\infty$ corresponds to the infinite capacity model. Thus we can use the results obtained in chapter 5 for the infinite capacity model. This leads to

$$(7.5.8)\qquad k_1(0;\infty) = \frac{1}{\sigma_1(s^*)^2},$$

$$(7.5.9)\qquad k_1(M-m;\infty) \cong \frac{1}{\lambda E[A]-\sigma_1}\left[\frac{(M-m)^2}{2} - \frac{E[Z_1^3]}{6E[Z_1]} + \frac{\lambda E[A^2]}{2(\lambda E[A]-\sigma_1)} \times \right.$$
$$\left. \times \left(M-m+\frac{E[Z_1^2]}{2E[Z_1]}\right)\right],$$

provided M-m satisfies condition 7.4.1. The probability distribution function of Z_1 is given by (7.4.26).

By applying the Key Renewal Theorem to the equation

$$\int_0^x G'(x-y)dR(y) = R'(x),$$

we obtain

$$(7.5.10)\qquad \lim_{M\to\infty} R'(M) = \frac{s^*}{1+\frac{\lambda}{\sigma_1}\tilde{F}'(s^*)}.$$

Incidentally, the result (7.5.8) can be derived from (7.5.10) and the application of the Key Renewal Theorem to (7.5.6).

To obtain an expression for $k_1(x;M)$ we combine (7.5.5), (7.5.7) and (7.5.8), yielding

$$(7.5.11)\qquad k_1(x;M) = k_1(x,\infty) + \left(k_1(0;M) - \frac{1}{\sigma_1(s^*)^2}\right)\int_0^x e^{s^*y}dR(y).$$

It remains to find approximate expressions for $k_1(0;M)$ and $\int_0^x e^{s^*y}dR(y)$. We first focus on the derivation of an approximation to the latter integral.

We recall the definition of $t_E(z;M)$, where we now explicitly express the dependence on M.

$t_E(z;M)$ = the expected amount of time that the dam is empty until the first overshoot of the level M, given that at epoch 0 the content is z and release rate σ_1 is used.

It follows from (7.4.7), (7.4.12) and (7.4.13) that

$$(7.5.12)\qquad t_E(z;M) = (\sigma_1 R'(M))^{-1} e^{-s^* z} \int_0^{M-z} e^{-s^*(M-z-y)} dR(y), \quad 0 \leq z \leq M.$$

Letting $M\to\infty$, we obtain by the Key Renewal Theorem and (7.5.10)

$$(7.5.13)\qquad t_E(z;\infty) = \frac{e^{-s^* z}}{\sigma_1 s^*}, \quad x \geq 0.$$

Next we express $t_E(z;M)$ in terms of $t_E(z;\infty)$. By conditioning on the overshoot $U(z;M)$ of level M we find

$$(7.5.14)\qquad t_E(z;\infty) = t_E(z;M) + \int_0^{\infty} t_E(M+u;\infty) dP\{U(z;M)\leq u\}.$$

Now it follows from the correspondence between the dam model and the production-inventory model and (1.3.26) that

$$(7.5.15)\qquad \lim_{M\to\infty} P\{U(z;M)\leq u\} = \frac{\lambda}{\lambda E[A]-\sigma_1} \int_u^{\infty} (1-e^{-s^*(y-u)})(1-F(y))dy.$$

A combination of (7.5.13)-(7.5.15) yields

$$(7.5.16)\qquad \lim_{M\to\infty} e^{s^* M} [t_E(z;M) - \frac{e^{-s^* z}}{\sigma_1 s^*}] = \frac{-(1+\frac{\lambda}{\sigma_1}\tilde{F}'(s^*))}{s^*(\lambda E[A]-\sigma_1)}$$

Note that (7.5.16) provides an alternative approximation to $t_E(z;M)$. Substitution of (7.5.12) into (7.5.16) and using (7.5.10) leads to

$$(7.5.17)\qquad \lim_{x\to\infty} \{ \int_0^x e^{s^* y} dR(y) - [\frac{e^{s^* x}}{1+\frac{\lambda}{\sigma_1}\tilde{F}'(s)} - \frac{\sigma_1}{\lambda E[A]-\sigma_1}]\} = 0.$$

Assuming M-m satisfies condition 7.4.1, we now find that

$$(7.5.18)\qquad \int_0^{M-m} e^{s^* y} dR(y) \cong \frac{e^{s^*(M-m)}}{1+\frac{\lambda}{\sigma_1}\tilde{F}'(s^*)} - \frac{\sigma_1}{\lambda E[A]-\sigma_1}.$$

This leaves us with the problem of finding a tractable expression for $k_1(0;M)$. We define for the production-inventory model with storage capacity M,

$v_1(x;M)$ = the expected amount of time until the inventory level drops below 0 for the first time, given that at epoch 0 the inventory equals x and production rate σ_1 is used.

Again we apply the arguments used to derive (7.4.12), yielding

$$(7.5.19)\qquad v_1(x;M) = v_1(0;M)\int_0^x e^{s^*y}dR(y) - \int_0^x \frac{(x-y)}{\sigma_1} e^{s^*y}dR(y)$$

$$(7.5.20)\qquad v_1(0;M) = (\sigma_1 R'(M))^{-1}\int_0^M e^{-s^*(M-y)}dR(y).$$

Analogously to (7.5.11) we have that

$$(7.5.21)\qquad v_1(x;M) = v_1(x;\infty) + (v_1(0;M) - \frac{1}{\sigma_1 s^*})\int_0^x e^{s^*y}dR(y).$$

It is obvious that $v_1(x;\infty)$ equals the function $t_1(x)$ which was defined in chapter 1 for the backlog model. Using (1.3.10) and (1.3.27) we find

$$(7.5.22)\qquad \lim_{x\to\infty}\{v_1(x;\infty) - [\frac{x}{\lambda E[A]-\sigma_1} + \frac{\lambda E[A^2]}{2(\lambda E[A]-\sigma_1)^2} - \frac{1}{s^*(\lambda E[A]-\sigma_1)}]\} = 0.$$

The next step is to combine equation (7.5.6), which gives an expression for $k_1(0;M)$, with equations (7.5.19)-(7.5.21). This yields

$$(7.5.23)\qquad e^{s^*M}k_1(0;M) = \frac{\int_0^M e^{s^*y}dR(y)}{\sigma_1 s^* R'(M)} - \frac{v_1(M;\infty)}{R'(M)}.$$

Then it follows from (7.5.10), (7.5.17), (7.5.22) and (7.5.23) that

$$\lim_{M\to\infty} e^{s^*M}\{k_1(0;M) - [\frac{1}{\sigma_1(s^*)^2} - \frac{(1+\frac{\lambda}{\sigma_1}\tilde{F}'(s^*))}{s^*}e^{-s^*M}\times$$

$$\times(\frac{M}{\lambda E[A]-\sigma_1} + \frac{\lambda E[A^2]}{2(\lambda E[A]-\sigma_1)^2})]\} = 0.$$

Hence, assuming M satisfies condition 7.4.1 we have

$$(7.5.24) \qquad k_1(0;M) \cong \frac{1}{\sigma_1(s^*)^2} - \frac{(1+\frac{\lambda}{\sigma_1}\tilde{F}'(s^*))}{s^*} e^{-s^*M} \times$$

$$\times \left(\frac{M}{\lambda E[A]-\sigma_1} + \frac{\lambda E[A^2]}{2(\lambda E[A]-\sigma_1)^2}\right).$$

Substitution of the approximate expressions (7.5.9), (7.5.18) and (7.5.24) into equation (7.5.11) leads to a tractable expression for $k_1(M-m;M)$. Then we use this expression together with (7.5.3) to obtain an approximation for the average content from equation (7.5.1). This approximation is tested numerically in the next section.

Remark 7.5.1. One might suggest that the approach outlined in this section is also applicable to obtain an alternative approximation to $p(m,u)$ for the case of $\sigma_1>0$. However, the approximation that follows from the arguments used to derive (7.5.11) leads to the approximate expression (7.4.27). This approximation is applicable when $c_A^2 \leqq 1$, but it does not lead to accurate results when $c_A^2>1$. Therefore it is necessary to apply the approximation $\hat{R}(x)$ to $R(x)$ given by (7.4.23).

Remark 7.5.2. In the chapters 1 to 6 we studied infinite capacity production-inventory models. By applying the approximations for $t_1(m)$ and $p(m,u)$ obtained in section 7.4 and the approximation for $k_1(x;M)$ obtained in this section, we obtain approximations to the service levels and costs in the finite capacity production-inventory models with $\pi_1>0$.

7.6. Numerical results and conclusions.

In this section we show that the approximations derived in this chapter lead to practically useful results. Using the results obtained in the sections 7.2-7.4 we derive (m,M)-rules that approximately satisfy one of the following service level constraints.

1. the number of upcrossings of level U per unit time equals $1-\alpha$.
2. the fraction of input lost equals $1-\beta$.
3. the fraction of time the dam is empty equals $1-\varepsilon$.

In all examples we take $\lambda=1$ and $E[A]=1$. The small release rate σ_1 has the two values 0.25 and 0.75. We do not consider the case of $\sigma_1 \leq 0$, since in that case the dam model is equivalent to the production-inventory models studied in the chapters 1 and 2 by noting that for $\sigma_1 \leq 0$ these models are in fact finite capacity models. The fast release rate π_2 has the two values 1.25 and 2. We assume that a switch-over cost of K=25 is incurred each time the release rate is switched from σ_1 to σ_2, while a holding cost at rate h.x is incurred when the content equals x, where h=1 is chosen. Then we predetermine M-m by

$$(7.6.1) \qquad M-m = \left\{\frac{2K(\sigma_2-\lambda E[A])(\lambda E[A]-\sigma_1)}{h(\sigma_2-\sigma_1)}\right\}^{\frac{1}{2}}$$

The formula (7.6.1) is derived along the same lines as the EOQ-formula (1.5.1). We consider four input distributions, whose squared coefficients of variation c_A^2 range from 0 to 2,

(i) deterministic input ($c_A^2=0$).
(ii) Erlang-2 input ($c_A^2=0.5$).
(iii) exponential input ($c_A^2=1$).
(iv) hyperexponential input ($c_A^2=2$) with F given by (7.4.32) and $p/\mu_1=(1-p)/\mu_2$ (balanced means).

Each of the required service levels α, β and ε is varied as 0.95 and 0.99.

In tables 7.6.1, 7.6.2 and 7.6.3 we give the approximate (m,M)-rules for the α-, β- and ε-service measure. The actual service levels α_{act}, β_{act} and ε_{act} are determined by computer simulation. In all examples we have simulated 250,000 inputs. Again the notation 0.950(3) means that the 95% confidence interval of the simulated value is given by 0.947-0.953.

The results from tables 7.6.1-7.6.3 show that the approximations derived in this chapter are quite accurate. For comments on the sensitivity of the switching level m to the underlying input distribution we refer to section 1.5.

Table 7.6.1. The approximate (m,M)-rules and their actual α-service levels.

C=∞			c_A^2=0				c_A^2=0.5			
σ_1	σ_2	α	U	m	M	α-act	U	m	M	α-act
0.25	1.25	0.95	8	1.52	4.58	0.951(3)	10	2.59	5.65	0.950(2)
0.25	2.00	0.95	8	2.30	6.93	0.949(2)	10	3.69	8.31	0.949(2)
0.75	1.25	0.95	8	2.73	5.23	0.951(2)	10	4.25	6.75	0.951(2)
0.75	2.00	0.95	8	4.11	7.28	0.951(2)	10	5.81	8.97	0.950(1)
0.25	1.25	0.99	12	1.78	4.84	0.991(2)	15	1.70	4.77	0.991(2)
0.25	2.00	0.99	12	4.99	9.62	0.990(1)	15	6.45	11.07	0.990(1)
0.75	1.25	0.99	12	3.01	5.51	0.990(2)	15	3.28	5.78	0.990(2)
0.75	2.00	0.99	12	6.85	10.01	0.990(1)	15	8.60	11.76	0.990(1)
C=U			c_A^2=0				c_A^2=0.5			
σ_1	σ_2	α	U	m	M	α-act	U	m	M	α-act
0.25	1.25	0.95	8	1.31	4.37	0.951(3)	10	2.09	5.16	0.950(3)
0.25	2.00	0.95	8	2.28	6.91	0.950(1)	10	3.64	8.27	0.950(2)
0.75	1.25	0.95	8	2.51	5.01	0.951(2)	10	3.71	6.21	0.951(2)
0.75	2.00	0.95	8	4.10	7.26	0.949(2)	10	5.76	8.92	0.951(2)
0.25	1.25	0.99	12	1.74	4.80	0.991(1)	15	1.60	4.66	0.991(2)
0.25	2.00	0.99	12	4.99	9.62	0.990(1)	15	6.44	11.06	0.990(1)
0.75	1.25	0.99	12	2.97	5.47	0.990(2)	15	3.16	5.66	0.990(1)
0.75	2.00	0.99	12	6.85	10.01	0.990(1)	15	8.59	11.75	0.990(1)
C=∞			c_A^2=1				c_A^2=2			
σ_1	σ_2	α	U	m	M	α-act	U	m	M	α-act
0.25	1.25	0.95	10	2.13	5.19	0.950	10	2.51	5.57	0.951(2)
0.25	2.00	0.95	10	3.25	7.87	0.950	10	2.90	7.53	0.951(2)
0.75	1.25	0.95	10	4.07	6.57	0.950	10	5.05	7.55	0.949(2)
0.75	2.00	0.95	10	5.61	8.77	0.950	10	5.69	8.85	0.950(2)
0.25	1.25	0.99	20	4.09	7.15	0.990	25	4.54	7.60	0.990(2)
0.25	2.00	0.99	20	10.03	14.66	0.990	25	12.07	16.70	0.989(1)
0.75	1.25	0.99	20	6.22	8.72	0.990	25	7.43	9.93	0.990(1)
0.75	2.00	0.99	20	12.52	15.68	0.990	25	15.32	18.48	0.990(1)
C=U			c_A^2=1				c_A^2=2			
σ_1	σ_2	α	U	m	M	α-act	U	m	M	α-act
0.25	1.25	0.95	10	1.22	4.28	0.950	10	0.18	3.24	0.950(2)
0.25	2.00	0.95	10	3.15	7.78	0.950	10	2.61	7.24	0.950(2)
0.75	1.25	0.95	10	2.98	5.48	0.950	10	1.95	4.45	0.951(2)
0.75	2.00	0.95	10	5.51	8.67	0.950	10	5.38	8.54	0.949(1)
0.25	1.25	0.99	20	3.89	6.95	0.990	25	4.04	7.10	0.990(1)
0.25	2.00	0.99	20	10.01	14.64	0.990	25	12.00	16.63	0.990(1)
0.75	1.25	0.99	20	6.01	8.51	0.990	25	6.86	9.36	0.990(2)
0.75	2.00	0.99	20	12.50	15.66	0.990	25	15.25	18.42	0.990(1)

Table 7.6.2. The approximate (m,M)-rules and their actual β-service levels.

C=U			$c_A^2=0$				$c_A^2=0.5$			
σ_1	σ_2	β	U	m	M	β-act	U	m	M	β-act
0.25	1.25	0.95	10	4.87	7.94	0.950(2)	10	3.07	6.13	0.950(2)
0.25	2.00	0.95	10	4.97	9.59	0.950(1)	10	4.12	8.75	0.950(2)
0.75	1.25	0.95	10	6.18	8.68	0.951(2)	10	4.76	7.26	0.950(2)
0.75	2.00	0.95	10	6.83	9.99	0.950(1)	10	6.25	9.41	0.949(2)
0.25	1.25	0.99	10	1.47	4.53	0.990(1)	15	2.70	5.77	0.991(1)
0.25	2.00	0.99	10	3.72	8.35	0.990(1)	15	6.94	11.57	0.990(1)
0.75	1.25	0.99	10	2.68	5.18	0.990(1)	15	4.37	6.87	0.991(1)
0.75	2.00	0.99	10	5.57	8.73	0.990(1)	15	9.09	12.26	0.990(1)
C=U			$c_A^2=1$				$c_A^2=2$			
σ_1	σ_2	β	U	m	M	β-act	U	m	M	β-act
0.25	1.25	0.95	10	1.22	4.28	0.950	15	2.19	5.25	0.951(2)
0.25	2.00	0.95	10	3.15	7.78	0.950	15	5.56	10.19	0.950(3)
0.75	1.25	0.95	10	2.98	5.48	0.950	15	4.69	7.19	0.950(3)
0.75	2.00	0.95	10	5.51	8.67	0.950	15	8.65	11.82	0.950(3)
0.25	1.25	0.99	20	3.89	6.95	0.990	25	0.29	3.35	0.989(2)
0.25	2.00	0.99	20	10.01	14.64	0.990	25	9.78	14.41	0.990(2)
0.75	1.25	0.99	20	6.01	8.51	0.990	25	2.10	4.60	0.990(2)
0.75	2.00	0.99	20	12.50	15.66	0.990	25	13.00	16.16	0.990(2)

In table 7.6.4 we show the accuracy of the approximation to the average content, which was derived in section 7.5. For the (m,M)-rules given in table 7.6.1 we have given V_{app} and V_{act}, respectively the approximate and simulated value of the average content. For the simulated value we have given also the 95% confidence interval, using the usual notation. For the case of exponential input the approximation to the average content V_{app} is exact.

Assuming linear holding costs and fixed switching costs, it is now possible to derive expressions for the average holding and switching costs. As in chapter 5 and 6 one can determine an approximate (m,M)-rule that satisfies some service level constraint with minimal average costs. We omit further details.

Table 7.6.3. The approximate (m,M)-rules and their actual ε-service levels.

C=∞			c_A^2=0				c_A^2=0.5			
σ_1	σ_2	ε	U	m	M	ε-act	U	m	M	ε-act
0.25	1.25	0.95	8	0.03	3.09	0.950(1)	10	0.04	3.11	0.949(1)
0.25	2.00	0.95	8	0.14	4.77	0.950(1)	10	0.18	4.81	0.950(1)
0.75	1.25	0.95	8	0.98	3.48	0.951(2)	10	1.69	4.19	0.952(2)
0.75	2.00	0.95	8	1.53	4.69	0.950(2)	10	2.53	5.70	0.950(2)
0.25	1.25	0.99	8	0.44	3.50	0.990(1)	10	0.51	3.57	0.990(1)
0.25	2.00	0.99	8	0.55	5.18	0.990(1)	10	0.64	5.27	0.990(1)
0.75	1.25	0.99	8	3.46	5.96	0.990(1)	10	5.17	7.67	0.990(1)
0.75	2.00	0.99	8	4.01	7.17	0.990(1)	10	6.02	9.18	0.989(2)
C=U			c_A^2=0				c_A^2=0.5			
σ_1	σ_2	ε	U	m	M	ε-act	U	m	M	ε-act
0.25	1.25	0.95	8	0.04	3.10	0.950(1)	10	0.07	3.13	0.950(1)
0.25	2.00	0.95	8	0.15	4.77	0.950(1)	10	0.18	4.81	0.949(1)
0.75	1.25	0.95	8	1.06	3.56	0.950(2)	10	1.86	4.36	0.952(2)
0.75	2.00	0.95	8	1.53	4.70	0.950(2)	10	2.54	5.70	0.951(2)
0.25	1.25	0.99	8	0.45	3.51	0.990(1)	10	0.53	3.59	0.990(1)
0.25	2.00	0.99	8	0.55	5.18	0.990(1)	10	0.64	5.27	0.990(1)
0.75	1.25	0.99	8	3.70	6.20	0.990(1)	10	5.71	8.21	0.990(1)
0.75	2.00	0.99	8	4.04	7.20	0.990(1)	10	6.12	9.28	0.991(1)
C=∞			c_A^2=1				c_A^2=2			
σ_1	σ_2	ε	U	m	M	ε-act	U	m	M	ε-act
0.25	1.25	0.95	10	0.07	3.14	0.950	10	0.07	3.13	0.949(2)
0.25	2.00	0.95	10	0.25	4.87	0.950	10	0.26	4.89	0.950(1)
0.75	1.25	0.95	10	2.39	4.89	0.950	10	3.29	5.79	0.953(2)
0.75	2.00	0.95	10	3.55	6.71	0.950	10	4.99	8.15	0.948(2)
0.25	1.25	0.99	15	0.61	3.67	0.990	15	0.65	3.71	0.990(1)
0.25	2.00	0.99	15	0.77	5.40	0.990	15	0.85	5.47	0.990(1)
0.75	1.25	0.99	15	6.88	9.38	0.990	15	9.44	11.94	0.990(2)
0.75	2.00	0.99	15	8.05	11.21	0.990	15	11.13	14.29	0.990(2)
C=U			c_A^2=1				c_A^2=2			
σ_1	σ_2	ε	U	m	M	ε-act	U	m	M	ε-act
0.25	1.25	0.95	10	0.12	3.18	0.950	10	0.17	3.23	0.950(2)
0.25	2.00	0.95	10	0.25	4.87	0.950	10	0.28	4.91	0.950(2)
0.75	1.25	0.95	10	2.86	5.36	0.950	10	4.63	7.13	0.949(2)
0.75	2.00	0.95	10	3.60	6.76	0.950	10	5.29	8.45	0.950(3)
0.25	1.25	0.99	15	0.63	3.69	0.990	15	0.71	3.77	0.990(1)
0.25	2.00	0.99	15	0.77	5.40	0.990	15	0.85	5.48	0.991(1)
0.75	1.25	0.99	15	7.30	9.80	0.990	15	11.11	13.61	0.991(2)
0.75	2.00	0.99	15	8.09	11.25	0.990	15	11.57	14.73	0.990(2)

Table 7.6.4. The approximate average contents and their actual values.

	C=∞	$c_A^2=0$				$c_A^2=\frac{1}{2}$			
σ_1	σ_2	U	m	V_{app}	V_{act}	U	m	V_{app}	V_{act}
0.25	1.25	8	1.52	4.60	4.59(4)	10	2.59	6.53	6.51(10)
0.25	2	8	2.30	4.66	4.66(2)	10	3.69	6.14	6.15(2)
0.75	1.25	8	2.73	4.46	4.46(5)	10	4.25	6.31	6.39(10)
0.75	2	8	4.11	4.56	4.56(3)	10	5.81	5.81	5.82(4)
0.25	1.25	12	1.78	4.86	4.85(5)	15	1.70	5.65	5.63(9)
0.25	2	12	4.99	7.35	7.36(2)	15	6.45	8.90	8.91(2)
0.75	1.25	12	3.01	4.70	4.67(6)	15	3.28	5.49	5.41(8)
0.75	2	12	6.85	7.16	7.15(4)	15	8.60	8.43	8.47(5)
	C=U	$c_A^2=0$				$c_A^2=\frac{1}{2}$			
σ_1	σ_2	U	m	V_{app}	V_{act}	U	m	V_{app}	V_{act}
0.25	1.25	8	1.31	3.84	3.84(2)	10	2.09	4.92	4.93(2)
0.25	2	8	2.28	4.55	4.56(1)	10	3.64	5.91	5.91(1)
0.75	1.25	8	2.51	3.71	3.71(3)	10	3.71	4.70	4.71(4)
0.75	2	8	4.10	4.46	4.46(3)	10	5.76	5.57	5.58(4)
0.25	1.25	12	1.74	4.64	4.64(3)	15	1.60	5.16	5.17(5)
0.25	2	12	4.99	7.33	7.34(1)	15	6.44	8.84	8.85(1)
0.75	1.25	12	2.97	4.48	4.47(3)	15	3.16	5.00	5.00(5)
0.75	2	12	6.85	7.14	7.13(4)	15	8.59	8.36	8.38(5)
	C=∞	$c_A^2=1$				$c_A^2=2$			
σ_1	σ_2	U	m	V_{app}	V_{act}	U	m	V_{app}	V_{act}
0.25	1.25	10	2.13	6.95	6.95	10	2.51	9.26	9.08(25)
0.25	2	10	3.25	5.82	5.82	10	2.90	5.85	5.83(4)
0.75	1.25	10	4.07	6.76	6.76	10	5.05	9.03	9.08(28)
0.75	2	10	5.61	5.43	5.43	10	5.69	5.43	5.41(6)
0.25	1.25	20	4.09	8.91	8.91	25	4.54	11.29	11.34(38)
0.25	2	20	10.03	12.60	12.60	25	12.07	15.02	15.08(4)
0.75	1.25	20	6.22	8.57	8.57	25	7.43	10.95	10.94(27)
0.75	2	20	12.52	11.82	11.82	25	15.32	14.02	13.98(10)
	C=U	$c_A^2=1$				$c_A^2=2$			
σ_1	σ_2	U	m	V_{app}	V_{act}	U	m	V_{app}	V_{act}
0.25	1.25	10	1.22	4.25	4.25	10	0.18	3.47	3.45(4)
0.25	2	10	3.15	5.40	5.40	10	2.61	4.82	4.82(2)
0.75	1.25	10	2.98	4.12	4.12	10	1.95	3.44	3.44(4)
0.75	2	10	5.51	5.01	5.01	10	5.38	4.44	4.45(4)
0.25	1.25	20	3.89	8.06	8.06	25	4.04	9.34	9.33(8)
0.25	2	20	10.01	12.48	12.48	25	12.00	14.70	14.69(3)
0.75	1.25	20	6.01	7.72	7.72	25	6.86	9.01	8.95(10)
0.75	2	20	12.50	11.70	11.70	25	15.25	13.68	13.67(11)

APPENDIX A. SOME RESULTS FOR A RANDOM WALK INDUCED BY A DISTRIBUTION FUNCTION WITH AN EXPONENTIAL TAIL.

In Feller [1971], p. 389-406, general results are derived concerning the distributions of ladder heights associated with a random walk in $\mathbb{R}$. Feller indicates how these results lead to exact expressions for these ladder height distributions when the random walk is induced by a distribution function with an exponential tail. For this particular type of random walk we provide more detailed information for those results that are needed in this monograph.

Let $\{D_n\}$ and $\{Q_n\}$ be two sequences of independent and identically distributed nonnegative random variables. The sequences $\{D_n\}$ and $\{Q_n\}$ are also independent of each other. We consider the particular case of

$$P\{D_n \leq x\} := F(x), \quad x \geq 0, \ n \geq 1,$$

$$P\{Q_n \leq y\} := 1-e^{-\mu y}, \quad y \geq 0, \ n \geq 1,$$

where F is a general probability distribution function concentrated on $[0,\infty)$. Define now the random walk $\{S_n\}$ by

$$S_0 := 0, \quad S_n := \sum_{i=1}^{n} X_i, \quad n \geq 1,$$

with

$$X_n = D_n - Q_n, \quad n \geq 1.$$

It is easily verified that

$$P\{X_1 \leq x\} = \begin{cases} \tilde{F}(\mu)e^{\mu x}, & x < 0 \\ e^{\mu x} \int_x^{\infty} F(z)\mu e^{-\mu z} dz, & x \geq 0, \end{cases} \tag{A.1}$$

with $\tilde{F}$ the Laplace-Stieltjes transform of F. Let

$$G(x) := P\{X_1 \leq x\}, \quad x \in \mathbb{R}.$$

We note that G(x) has an exponential tail. This is crucial in the discussion to follow.

We define the descending ladder points (W_k,τ_k) by

$$\tau_0 := 0,\ \tau_k := \min\{n \mid S_n<S_{\tau_{k-1}},\ \tau_n>\tau_{n-1}\},\quad k \geq 1.$$

$$W_k := S_{\tau_k},\quad k \geq 0.$$

The random variable W_k is the k-th descending ladder height and τ_k is the k-th descending ladder epoch. Note that X_1 has a continuous distribution function. Thus there is no ambiguity in the definition of the ladder points; the weak descending ladder points are at the same time the strict descending ladder points. Similarly, we define the ascending ladder points (Z_k,σ_k), $k\geq 1$, by

$$\sigma_0 := 0,\ \sigma_k := \min\{n \mid S_n>S_{\sigma_{k-1}},\ \sigma_n>\sigma_{k-1}\},\quad k \geq 1.$$

$$Z_k := S_{\sigma_k},\quad k \geq 0.$$

We further define

$$H(x) := P\{Z_1 \leq x\},\quad x \geq 0.$$

$$\rho(y) := P\{W_1 \leq y\},\quad y \leq 0.$$

$$\psi(x) := \sum_{k=0}^{\infty} P\{Z_k \leq x\},\quad x \geq 0.$$

$$\phi(y) := \sum_{k=0}^{\infty} P\{W_k \geq y\},\quad y \leq 0.$$

The functions $\psi(x)$ and $\phi(y)$ denote respectively the expected number of ascending ladder heights in $[0,x]$ and the expected number of descending ladder heights in $[y,0]$.

Throughout this appendix we assume that $E[X_1]>0$. Then it follows from Wald's equation that

$$E[Z_1] = E[\sigma_1]E[X_1]. \tag{A.2}$$

Also, the sequences $\{Z_k\}$ and $\{\sigma_k\}$ constitute proper renewal processes, whereas $\{W_k\}$ and $\{\tau_k\}$ constitute terminating renewal processes. Hence

(A.3) $\quad H(\infty) = 1$ and $\rho(0) < 1$.

Proceeding as in Feller [1971], p. 398-400, we obtain the following general relations

$$(A.4) \qquad H(x) = G(x)-G(0)-\int_{-\infty}^{0^-}[G(x-z)-G(-z)]d\phi(z), \qquad x \geqq 0.$$

$$(A.5) \qquad \rho(y) = \int_{0^-}^{\infty} G(y-z)d\psi(z), \qquad y \leqq 0.$$

$$(A.6) \qquad \psi(x) = 1-\rho(0)+\int_{0^-}^{\infty} G(x-z)d\psi(z), \qquad x \geqq 0.$$

$$(A.7) \qquad \phi(y) = 1-H(\infty)+1-G(y)-\int_{-\infty}^{0^-}[1-G(y-z)]d\phi(z), \qquad y \leqq 0.$$

It can be shown that these four equations uniquely determine the functions $H(x)$, $\rho(y)$, $\psi(x)$ and $\phi(y)$. These relations are the starting point for the exact computation of $H(x)$, $\rho(y)$ and $\phi(y)$ for the particular random walk considered.

Substituting (A.1) into (A.5) we obtain

$$(A.8) \qquad \rho(y) = \int_{0^-}^{\infty} e^{-\mu z}d\psi(z).\tilde{F}(\mu)e^{\mu y}, \qquad y \leqq 0.$$

From general results from renewal theory it follows that the integral in (A.8) is finite. Since the renewal process $\{W_k\}$ is terminating the probability distribution function $\rho(y)$ is defective and there exists some s^* with $0<s^*<\mu$, such that

$$(A.9) \qquad \rho'(y) = (\mu-s^*)e^{\mu y}, \qquad y \leqq 0,$$

with $\rho'(y)$ the density of $\rho(y)$. Using the definition of $\phi(y)$ and the fact that $\phi(y)$ has a density $\phi'(y)$, we obtain

$$(A.10) \qquad \phi'(y) = -(\mu-s^*)e^{s^* y}, \qquad y \leqq 0.$$

Substituting this into (A.4) and using (A.1), we obtain

$$H(x) = \int_0^\infty e^{-s^*\omega}\mu[F(x+\omega)-F(\omega)]d\omega. \tag{A.11}$$

Next we use (A.3) and (A.11) to find

$$1 = \mu \int_0^\infty e^{-s^*\omega}[1-F(\omega)]d\omega,$$

which is equivalent to

$$s^* = \mu(1-\tilde{F}(s^*)). \tag{A.12}$$

From the transcendental equation (A.12) we determine s^*. Knowing the quantity s^* we obtain the following exact expressions.

$$\rho(y) = \frac{(\mu-s^*)}{\mu} e^{\mu y}, \quad y \leq 0. \tag{A.13}$$

$$\phi(y) = 1 + \frac{(\mu-s^*)}{s^*}(1-e^{s^*y}), \quad y \leq 0. \tag{A.14}$$

$$1-H(x) = \mu \int_0^\infty e^{-s^*z}(1-F(x+z))dz, \quad x \geq 0. \tag{A.15}$$

$$E[\sigma_1] = \frac{\mu}{s^*} \tag{A.16}$$

The results obtained in this appendix have been applied in section 1.3 with $\mu=\lambda/\pi_1$. A similar approach based on the relations (A.4)-(A.7) can be used for an alternative derivation of the expression for q(x) in section 1.4. However, where the equation (A.12) has always a unique positive solution, we then obtain a transcendental equation that has a solution (being necessarily unique) only if the distribution function F has an exponentially decreasing tail.

REFERENCES.

[1] ALI KHAN, M.S., *Infinite dams with additive inputs*, J. Appl. Probab. 14 (1977) p. 170-180.

[2] ATTIA, F.A. and BROCKWELL, P.J., *The control of a finite dam*, J. Appl. Probab. 19 (1982) p. 815-825.

[3] BATHER, J.A., *A continuous time inventory model*, J. Appl. Probab. 3 (1966) p. 538-549.

[4] BROCKWELL, P.J., *Stationary distributions for dams with additive inputs and content-dependent release rate*, Adv. in Appl. Probab. 9 (1977) p. 645-663.

[5] BROCKWELL, P.J. and CHUNG, K.L., *Emptiness times of a dam with stable input and general release rule*, J. Appl. Probab. 12 (1975) p. 212-217.

[6] CINLAR, E., *Introduction to stochastic processes*, Prentice-Hall Englewood Cliffs, New Jersey (1975).

[7] COHEN, J.W., *On regenerative processes in queueing theory*, Lecture notes in economics and mathematical systems, Springer-Verlag Berlin 121 (1976).

[8] COX, D.R., *A use of complex probabilities in the theory of stochastic processes*, Proc. Camb. Phil. Soc. 51 (1955) p. 313-319.

[9] DE KOK, A.G., *Approximations for a lost-sales production/inventory control model with service level constraints*, to appear in Management Sci. (1985).

[10] DE KOK, A.G. and TIJMS, H.C., *A two-moments approximation for a buffer design problem requiring a small rejection probability*, to appear in Performance Evaluation (1985a).

[11] DE KOK, A.G. and TIJMS, H.C., *A stochastic production/inventory system with all-or-nothing demand and service measures*, to appear in Stochastic Models (1985b).

[12] DE KOK, A.G. and TIJMS, H.C., *A queueing system with impatient customers*, to appear in J. Appl. Probab. (1985c).

[13] DE KOK, A.G., TIJMS, H.C. and VAN DER DUYN SCHOUTEN, F.A., *Approximations for the single product production-inventory model with compound Poisson demand and service level constraints*, Adv. in Appl. Probab. 16 (1984) p. 378-401.

[14] DE KOK, A.G., TIJMS, H.C. and VAN DER DUYN SCHOUTEN, F.A., *A practical algorithm for the one product production/inventory problem*, to appear in European J. Oper. Res. (1985).

[15] DE LEVE, G., TIJMS, H.C. and WEEDA, P.J., *Generalized Markovian decision processes*, Math. Centre Tracts Mathematical Centre Amsterdam 5 (1976).

[16] DOSHI, B.T., *Two-mode control of a Brownian motion with quadratic loss and switching costs*, Stochastic Process. Appl. 6 (1978) p. 277-289.

[17] DOSHI, B.T., VAN DER DUYN SCHOUTEN, F.A. and TALMAN, A.J.J., *A production-inventory control model with a mixture of back-orders and lost-sales*, Management Sci. 24 (1978) p. 1078-1086.

[18] FADDY, M.J., *Optimal control of finite dams; discrete (2-stage) output procedure*, J. Appl. Probab. 11 (1974) p. 111-121.

[19] FELLER, W., *An introduction to probability theory and its applications*, Wiley New York 2, 2nd ed. (1971).

[20] GAVER, D.P., Jr., *Operating characteristics of a simple production/inventory problem under continuous review policy*, Oper. Res. 9 (1961) p. 635-649.

[21] GAVISH, B. and GRAVES, S.C., *A one-product production/inventory problem under continuous review policy*, Oper. Res. 28 (1980) p. 1228-1236.

[22] GRAVES, S.C., *Application of queueing theory to continuous perishable inventory systems*, Management Sci. 28 (1982) p. 400-404.

[23] GRAVES, S.C. and KEILSON, J., *The compensation method applied to a one-product production/inventory problem*, Math. Oper. Res. 6 (1981) p. 246-262.

[24] HADLEY, G. and WHITIN, T.M., *Analysis of inventory systems*, Prentice-Hall, Englewood Cliffs, New Jersey (1963).

[25] MORAN, P.A.P., *A theory of dams with continuous input and a general release rule*, J. Appl. Probab. 6 (1969) p. 88-98.

[26] PETERSON, R. and SILVER, E.A., *Decision systems for inventory management and production planning*, Wiley New York (1979).

[27] PUTERMAN, M., *A diffusion process model for a storage system*, Logistics, North-Holland/TIMS studies in the Management Sciences, ed. by M. Geisler, North-Holland Amsterdam 1 (1975) p. 143-159.

[28] ROSS, S.M., *Applied probability models with optimization applications*, Holden-Day San Francisco (1970).

[29] SCHASSBERGER, R., *Warteschlangen*, Springer-Verlag Wien-New York (1973).

[30] SCHNEIDER, H., *Effect of service-levels on order-points or order-levels in inventory models*, Internat. J. Production Res. 19 (1981) p. 615-631.

[31] SMITH, N.M.H. and YEO, G.F., *On a general storage problemand and its approximate solution*, Adv. in Appl. Probab. 13 (1981) p. 567-602.

[32] SOBEL, M.J., *Optimal average cost policy for a queue with start-up and shut-down costs*, Oper. Res. 18 (1970) p. 145-162.

[33] STIDHAM, S., Jr., *Regenerative processes in the theory of queues with applications to the alternating priority queue*, Adv. in Appl. Probab. 4 (1972) p. 542-577.

[34] TIJMS, H.C., *On a switch-over policy for controlling the workload in a queueing system with two constant service rates and fixed switch-over costs*, Z. Oper. Res. 21 (1977) p. 19-32.

[35] TIJMS, H.C., *An algorithm for average costs denumerable state semi-Markov decision problems with applications to controlled production and queueing systems*, Recent developments in Markov decision processes, ed. by R. Hartley, L.C. Thomas and D.J. White Academic Press New York (1980) p. 143-180.

[36] TIJMS, H.C., *Stochastic operations research: a computational approach*, Wiley Chichester (1986).

[37] TIJMS, H.C. and GROENEVELT, H., *Simple approximations for the reorder point in periodic and continuous review (s,S) inventory systems with service level constraints*, European J. Oper. Res. 17 (1984) p. 175-190.

[38] TIJMS, H.C. and VAN DER DUYN SCHOUTEN, F.A., *Inventory control with two switch-over levels for a class of M/G/1 queueing systems with variable arrival and service rate*, Stochastic Process. Appl. 6 (1978) p. 213-222.

[39] VICKSON, R.G., *A single-product cycling problem under Brownian motion demand*, Report of Dept. of Manag. Sci. University of Waterloo Ontario (1982).

[40] WHITT, W., *Approximating a point process by a renewal process, 1: two basic methods*, Oper. Res. 30 (1982) p. 125-147.

[41] WHITT, W., *On approximations for queues, 1: extremal distributions*, Bell System Tech. J. 63 (1984) p. 115-138.

[42] WOLFF, R.W., *Poisson arrivals see time averages*, Oper. Res. 30 (1982) p. 223-231.

[43] YEO, G.F., *A finite dam with variable release rate*, J. Appl. Probab. 12 (1975) p. 205-211.

[44] ZUCKERMAN, D., *Two-stage output procedure of a finite dam*, J. Appl. Probab. 14 (1977) p. 421-425.

SAMENVATTING.

Dit proefschrift behandelt de probabilistische analyse van een aantal één-produkt produktie-voorraadmodellen met een samengesteld Poisson vraagproces en twee mogelijke produktiesnelheden.

Wordt in klassieke voorraadmodellen de voorraad aangevuld door bestellingen, in de hier geanalyseerde produktie-voorraadmodellen gebeurt dat door continue produktie. Centraal staat het vraagstuk, hoe de produktiesnelheid met het voorraadniveau te coördineren, opdat de stochastische schommelingen in de vraag op adequate wijze worden opgevangen. Hierbij verstaan we onder "op adequate wijze" dat de produktiesnelheid zo wordt bestuurd dat aan een of ander servicecriterium wordt voldaan. Voorbeelden van algemeen gebruikte servicematen zijn het gemiddeld aantal keren per tijdseenheid, dat het systeem buiten voorraad raakt en de fraktie van de vraag, waaraan direkt uit voorraad kan worden voldaan. Doel van de analyse is het verkrijgen van praktisch bruikbare beslisregels. De voornaamste middelen die we hiervoor gebruiken, zijn resultaten uit de theorie van de één-dimensionale stochastische wandelingen en asymptotische resultaten uit de vernieuwingstheorie.

Bij de keuze van een beslisregel, die vastlegt hoe snel geproduceerd wordt op ieder tijdstip, richt men zich meestal op de minimalisatie van de gemiddelde kosten per tijdseenheid. We onderscheiden omstelkosten voor het omschakelen van de produktiesnelheid, voorraadkosten, produktiekosten en kosten voor tekorten. Terwijl de eerste drie kostensoorten doorgaans eenvoudig te specificeren zijn, is dit niet het geval bij kosten voor tekorten. Immers, hoe kan het verlies van goodwill gequantificeerd worden; wat zijn de kosten van toekomstige verliezen en omzetdaling wanneer niet of niet direkt aan de vraag van een klant kan worden voldaan? Daarom worden in de praktijk de kosten voor tekorten indirekt bepaald door het gebruik van servicematen. Men eist bijvoorbeeld dat tenminste 95% van de klanten direkt moet worden geholpen.

Een beslisregel die bepaalt dat de produktiesnelheid altijd hoger is dan de gemiddelde vraag per tijdseenheid zal hoge voorraadkosten veroorzaken. Echter, indien de produktiesnelheid altijd lager is dan de gemiddelde vraag per tijdseenheid zal de service aan klanten tekort schieten. Een beslisregel die een compromis biedt tussen deze twee uitersten en daarbij ook gemakkelijk implementeerbaar is, is de zgn. (m,M)-regel. Onder deze beslisregel wordt de produktiesnelheid van hoog naar laag omgeschakeld, zodra het voorraadniveau tenminste M bedraagt. De produktiesnelheid wordt

weer omgeschakeld van laag naar hoog, zodra het voorraadniveau lager wordt dan m. In dit proefschrift zullen we steeds aannemen dat de produktie, en dus de voorraad, wordt bestuurd door een (m,M)-regel.

Onze aanpak is als volgt. Allereerst leiden we voor een gegeven (m,M)-regel eenvoudige approximaties af voor de serviceniveaus behorend bij een aantal veel toegepaste servicematen. Louter rekening houdend met voorraad- en omstelkosten bepalen we dan het verschil M-m. Tenslotte bepalen we m zo, dat aan een bepaalde service-eis wordt voldaan. Met behulp van computersimulatie tonen we vervolgens aan dat de verkregen approximaties zeer accuraat zijn. Daarom kunnen de approximaties ook gebruikt worden voor gevoeligheidsanalyses.

Het bepalen van een (m,M)-regel, die voorraad- en omstelkosten minimaliseert onder een servicerestrictie is in principe een 2-dimensionaal probleem. Door de hierboven beschreven sequentiële bepaling van M-m en m besparen we veel rekenwerk. Aan de hand van numerieke resultaten laten we zien dat de eenvoudige formules, die we hanteren voor de bepaling van M-m, beslisregels opleveren met lage kosten. Deze formules zijn niet alleen onafhankelijk van de servicemaat, maar bovendien onafhankelijk van de manier waarop een tekort veroorzaakt door een vraag die groter is dan de aanwezige voorraad wordt behandeld (bijv. nalevering of geen nalevering).

In hoofdstuk 1 beschouwen we het produktie-voorraadmodel, waarin tekorten worden nageleverd (het "backlog" model). Gebruikmakend van asymptotische resultaten uit de vernieuwingstheorie en (in appendix A beschreven) resultaten voor stochastische wandelingen gegenereerd door een verdelingsfunktie met exponentiële staart, vinden we approximaties voor de serviceniveaus van een aantal servicematen bij een gegeven (m,M)-regel. Numerieke resultaten tonen de accuratesse van deze benaderingen aan. Voor een gegeven servicerestrictie en vast verschil M-m bestuderen we de gevoeligheid van het omstelniveau m voor derde en hogere momenten van de vraagverdeling.

In hoofdstuk 2 nemen we aan dat elk tekort verloren gaat (het "lost-sales" model). Tussen het "lost-sales" model en het "backlog" model worden exacte relaties afgeleid, die resulteren in approximaties voor het lost-sales model. Naast het aantonen van de praktische bruikbaarheid van deze approximaties gaan we ook na of we op grond van de eerste twee momenten van de vraag per tijdseenheid, zonder de waarde van de aankomstrate λ expliciet te kennen, een (m,M)-regel kunnen bepalen, die voldoet aan een vooraf opgegeven service-eis.

In het model bestudeerd in hoofdstuk 3 mag vraag groter dan de fysieke voorraad tot een zekere hoeveelheid worden nageleverd. Alle vraag die leidt tot een grotere achterstand gaat verloren.

In hoofdstuk 4 beschouwen we het produktie-voorraadmodel waarin klanten die bij aankomst een achterstand aantreffen groter dan een gegeven vaste hoeveelheid, onmiddellijk vertrekken. Is de achterstand kleiner dan de gegeven vaste hoeveelheid, dan wordt aan hun vraag (al of niet direkt) voldaan.

Zowel in hoofdstuk 3 als in hoofdstuk 4 leiden we weer exacte relaties af tussen het betreffende model en de modellen bestudeerd in voorgaande hoofdstukken. Op die manier vinden we benaderingen voor een groot aantal relevante servicematen.

In hoofdstuk 5 geven we goede benaderingen voor de gemiddelde voorraad- en omstelkosten per tijdseenheid, geldend voor alle modellen behandeld in de hoofdstukken 1 t/m 4. We bepalen een bij benadering gemiddeld kosten optimale (m,M)-regel, die voldoet aan een gegeven service-eis. Aan de hand van numerieke resultaten laten we zien dat de (m,M)-regel met M-m gekozen volgens een uitbreiding van de formule van Camp in kosten ten hoogste 5% boven de minimale kosten ligt. Bij een voldoend hoge servicegraad blijkt de optimale waarde van M-m een vaste waarde $\tilde{\Delta}^*$ aan te nemen. Deze $\tilde{\Delta}^*$ kan vooraf worden berekend op grond van de voorraad- en omstelkosten en blijkt onafhankelijk te zijn van de servicemaat en de wijze waarop tekorten worden behandeld. Kiezen we M-m gelijk aan $\tilde{\Delta}^*$ dan blijken de kosten van de resulterende (m,M)-regel ten hoogste 1% boven de minimale kosten te liggen.

In hoofdstuk 6 behandelen we opnieuw het backlog model en het lost-sales model. We nemen nu echter aan dat de lage produktiesnelheid 0 is en er een positieve omsteltijd nodig is om de produktie te starten. Voor beide modellen geven we weer approximaties voor diverse servicematen. Ook voor deze modellen blijkt dat een sequentiële bepaling van M-m en m bij een goede keuze van M-m leidt tot een (m,M)-regel met lage kosten.

In de hoofdstukken 1 t/m 6 nemen we steeds aan dat de opslagcapaciteit oneindig is. De resultaten uit hoofdstuk 7 maken het mogelijk op analoge wijze de corresponderende eindige-capaciteitsmodellen te analyseren. In hoofdstuk 7 beschouwen we een dammodel, waarin de inhoud van de dam kan afnemen met een hoge of een lage leegloopsnelheid. Het proces dat de toevoegingen aan de daminhoud beschrijft is een samengesteld Poisson proces. Voor een aantal servicematen geven we approximaties voor hun serviceniveaus. Ook geven we een benadering voor de gemiddelde inhoud van de dam.

CURRICULUM VITAE

De schrijver van dit proefschrift werd op 10 juni 1958 geboren te 's-Gravenhage. Na het eindexamen Atheneum-B aan het Thomas More College te 's-Gravenhage begon hij in september 1976 met de studie wiskunde aan de Rijksuniversiteit te Leiden. September 1978 werd het kandidaatsexamen afgelegd met hoofdvakken wiskunde en natuurkunde en bijvak sterrenkunde. Gedurende de doctoraal-fase volgde hij colleges bij de hoogleraren dr. J. Fabius, dr. A. Hordijk, dr.ir. L.A. Peletier, dr. B.M.S. van Praag, dr. M.N. Spijker, dr. A.C. Zaanen en dr. W.R. van Zwet. Onder leiding van de laatste schreef hij zijn afstudeerscriptie op het gebied van de mathematische statistiek. Op 27 februari 1981 legde hij cum laude het doctoraalexamen af met hoofdvak wiskunde en bijvak economie.

Van 1 september 1978 tot 1 maart 1981 was hij als student-assistent in dienst bij het Instituut voor Toegepaste Wiskunde en Informatica van de Rijksuniversiteit te Leiden. De werkzaamheden bestonden uit het geven van onderwijs aan kandidaats-studenten op het gebied van de waarschijnlijkheids-rekening, discrete wiskunde en informatica.

Van 1 april 1981 tot 31 maart 1985 werkte hij als wetenschappelijk medewerker in tijdelijke dienst bij de Interfaculteit der Actuariële Wetenschappen en Econometrie van de Vrije Universiteit te Amsterdam. Het onderzoek dat hij verrichtte onder leiding van prof.dr. H.C. Tijms op het gebied van de mathematische besliskunde heeft geleid tot dit proefschrift.